

Extensive spontaneous genome reduction in *Paraburkholderia sabiae*

Received: 9 February 2026

Accepted: 13 April 2026

Published online: 29 April 2026

Cite this article as: Bolli K., Waichler C., Liu Y. *et al.* Extensive spontaneous genome reduction in *Paraburkholderia sabiae*. *Sci Rep* (2026). <https://doi.org/10.1038/s41598-026-49026-9>

Kim Bolli, Colin Waichler, Yilei Liu, Daphné Golaz, Adam M. Deutschbauer, Leo Eberl, Sebastian J. Hug, Marta Torres & Gabriella Pessi

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

Extensive spontaneous genome reduction in *Paraburkholderia sabiae*

**Kim Bolli¹, Colin Waichler¹, Yilei Liu¹, Daphné Golaz¹, Adam M. Deutschbauer^{2,3}, Leo Eberl¹,
Sebastian J. Hug¹, Marta Torres^{1*} & Gabriella Pessi^{1*}**

¹Department of Plant and Microbial Biology, University of Zurich, CH-8057 Zurich, CH

²Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory,
Berkeley, CA 94720, USA

³Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA

***Correspondence:**

Gabriella Pessi, gabriella.pessi@botinst.uzh.ch

Marta Torres, marta.torres@botinst.uzh.ch

Keywords: rhizobium, secretion systems, large megaplasmid, chromosome, killing

Abstract

Paraburkholderia sabiae LMG24235^T is a beta-proteobacterium originally isolated from *Mimosa caesalpiniiifolia* nodules that efficiently outcompetes phytopathogens both *in vitro* and *in planta*. Here we show that *P. sabiae* LMG24235^T can spontaneously lose a 2.3-Mbp replicon, resulting in a reduction of genome size by approximately one quarter and representing the first reported case of a bacterium losing 23% of its genome in a single event. Phylogenetic analysis of the replication initiation protein A (RepA) revealed that this replicon is closely related to megaplasms found in other *Paraburkholderia* strains and a randomly barcoded transposon mutagenesis (RB-TnSeq) analysis showed that it contains very few predicted likely-essential genes, all of which are dispensable for growth. In contrast to the wild-type strain, *P. sabiae* lacking the second replicon ($\Delta 2$) exhibited a growth defect on D-trehalose and L-leucine, reduced motility and decreased survival under low-oxygen conditions. While growth in rich medium was otherwise unaffected, *P. sabiae* $\Delta 2$ was outcompeted by the wild-type *in vitro*, and *M. caesalpiniiifolia* nodules occupied by $\Delta 2$ fixed 40% less nitrogen compared to wild-type nodules. These results suggest that the second replicon is not an essential chromosomal element but rather a dispensable large megaplasmid that confers a competitive advantage to *P. sabiae* LMG24235^T in its natural environment.

Introduction

Paraburkholderia is one of several genera in the family *Burkholderiaceae*, which is well recognized for its high metabolic versatility and capability to thrive in different environmental niches. Established in 2014, this genus within the Pseudomonadota phylum comprises more than 95 species isolated from soil, water, fungi, and plants [1]. *Paraburkholderia* species have considerable agricultural and biotechnological potential, since they display notable features including the ability to promote plant growth, degrade complex organic compounds, and inhibit the growth of phytopathogens by producing antimicrobial agents [2]. They can colonize the endophytic tissues of crops such as tomato, maize, potato, and rice, particularly the root tissues. They are known for their potential to fix nitrogen [3–16], improve nutrient uptake [2] and produce phytohormones [17]. Most *Paraburkholderia* species possess large genomes compared to other bacteria, generally exceeding 7 megabase pairs (Mbp), with many strains ranging from approximately 8 to more than 11 Mbp. Examples include *Paraburkholderia phymatum* STM815^T (8.7 Mbp), *Paraburkholderia xenovorans* LB400 (9.7 Mbp), and *Paraburkholderia caribensis* MWAP64 (9.03 Mbp). *Paraburkholderia* genomes are often organized into multiple replicons, including several chromosomes and plasmids [18–20].

Some *Paraburkholderia* members can fix atmospheric nitrogen through symbiotic association with legumes and hence are classified as beta-rhizobia. The genes required to establish symbiosis with plants are usually localized on a dedicated plasmid. In rhizobia belonging to the Alphaproteobacteria (i.e. alpha-rhizobia), it was shown that these symbiotic plasmids can be lost when the strains are grown in non-selective environments, such as an *in vitro* culture [21, 22]. Most rhizobial *Paraburkholderia* species (e.g. *Paraburkholderia mimosarum*, *Paraburkholderia tuberum*, and *P. phymatum*) were isolated from nodules of *Mimosa* plants in South America and Southeast Asia [23–26], but members of this group have also been isolated from legumes native to the South African Fynbos [12, 27–29]. Importantly, beta-rhizobia consistently outcompete alpha-rhizobia in nodulating *Mimosa* species [11, 13, 30], a dominance shaped by genetic and environmental factors. *Paraburkholderia sabiae* LMG24235^T was isolated from nodules of *Mimosa caesalpinifolia* in Brazil [8] and has since been shown to inhibit the growth of important phytopathogens, such as *Pseudomonas syringae*, *Burkholderia gladioli*, *Ralstonia solanacearum* and *Pectobacterium carotovorum* [31]. Additionally, *P. sabiae* was observed to protect potato tubers from bacterial soft rot disease caused by *P. carotovorum* [31]. Sequencing of *P. sabiae* genome revealed the presence of four replicons, provisionally classified as chromosomes 1 (6.58 Mbp), chromosome 2 (2.3 Mbp), plasmid pSymb (0.615 Mbp) containing the symbiotic genes, and plasmid 2 (0.399 Mbp). A type VI secretion system (T6SS) on chromosome 2, T6SS-1, was previously demonstrated to be essential for outcompeting bacterial phytopathogens [31]. In the present work, we characterized a *P. sabiae* LMG24235^T derivative that spontaneously lost its ability to kill pathogens *in vitro*.

Genome sequencing revealed the complete loss of *P. sabiae*'s 2.3-Mbp replicon, suggesting that this replicon is a megaplasmid rather than a true essential chromosome. This assessment is additionally supported by analyses examining the replicon's replication and partitioning features and essential gene content. We further investigated the functional consequences of this large-scale genome reduction, showing that the megaplasmid contributes to metabolic versatility, motility, competitiveness, and symbiotic nitrogen fixation. Overall, this study reports a striking large-scale spontaneous genome reduction event and examines the impact of this loss on bacterial physiology and symbiosis, suggesting that the deleted genes are either redundant or conditionally beneficial, with loss resulting in a more streamlined but less versatile genome.

Materials and Methods

Bacterial strains, media, and culture conditions.

All bacteria strains, plasmids, oligonucleotides and primers are listed in Supplementary Table 1. If not stated otherwise, *Paraburkholderia* strains were grown aerobically at 28°C either stationary or shaking at 180 rpm in LB without salt (LBNS) or AB minimal medium [32] supplemented with 15 mM potassium succinate as a carbon source (ABS). For antibiotic-resistant *Paraburkholderia* strains, chloramphenicol (Cm) was used at 80 µg/mL, gentamicin (Gm) at 20 µg/mL, kanamycin (Km) at 10-50 µg/mL, polymyxin B (pB) at 50 µg/mL, and carbenicillin (Cb) at 50 µg/mL. *Pseudomonas syringae* DC3000 and *Pectobacterium carotovorum* were grown in LBNS with Gm 20 µg/mL. *Escherichia coli* was grown on LB with Cm 20 µg/mL, Gm 10 µg/mL, Cb at 50 µg/mL or Km 25 µg/mL. When needed, diaminopimelic acid (DAP) was added at a final concentration of 300 µM. The spontaneous $\Delta 2$ mutant was isolated by picking a single colony from a freshly streaked bacterial plate prepared from the original glycerol stock of *P. sabiae* (see first Results section entitled "Isolation of a *Paraburkholderia sabiae* LMG24235 strain unable to kill phytopathogens that has lost the 2.3 Mbp megaplasmid").

Construction of a GFP-tagged *P. sabiae* strain

A *kanR-gfp* construct was inserted between loci QEN71_RS00040 and QEN71_RS00045 (Supplementary Table 1) to generate WT-GFP. Gibson assembly was used to clone *kanR* (p34E-KmR) and *gfp* (pSHAFT-GFP) into pSHAFT2 linearized with XhoI [33]. Assembly reactions were performed with a molar ratio of 1 backbone to 7 of each insert fragment, and the reaction product was transformed into *E. coli* c118 λ -pir [34]. The resulting plasmid pSHAFT2_GFP_*sabiae* was sequenced using primers pSHAFT_F and pSHAFT_R1 (Supplementary Table 1). After triparental mating using *E. coli* c118 λ -pir carrying pSHAFT2_GFP_*sabiae* as donor strain, *E. coli* DH5a carrying a pRK2013 (Invitrogen) as helper strain and *P. sabiae* wild-type as recipient, *P. sabiae* fluorescent colonies were selected on ABS Cm80 Km50 plates to identify deletion mutants with a *kanR-gfp* insert (WT-GFP). A clone of this strain

that spontaneously lost the second replicon was identified (*P. sabiae* $\Delta 2$ -GFP). The genome of WT-GFP was sequenced using a MiSeq i100 system (Illumina, Inc, CA, USA).

Construction of a *P. sabiae* randomly barcoded transposon mutant library (RB-TnSeq)

The *P. sabiae* RB-TnSeq library was constructed via conjugation with *E. coli* WM3064 harboring the pHLL250 *mariner* transposon vector library (strain AMD290) [35]. Briefly, *P. sabiae* was grown overnight in 50 mL of LBNS at 28°C. The next morning, a 2-mL freezer stock of strain AMD290 (DAP auxotroph) was recovered in 50 mL LB with Cb50 and DAP at 37°C. When the OD₆₀₀ of the *E. coli* donor strain reached 1, 20 mL of the culture were harvested and washed three times with LB DAP medium. Then, 20 mL of *P. sabiae* cells (OD₆₀₀=1) were harvested, mixed with the washed donor cells, centrifuged, and resuspended in 0.5 mL LBNS DAP. The resuspension was spotted onto 0.45- μ m membrane filters (Millipore Sigma, MA, USA) and incubated at 28°C on LBNS DAP plates supplemented with 1.5% agar. The next day, the conjugation was scraped from the membrane, resuspended in 10 mL LBNS Km10 and plated at different dilutions on LBNS Km10 plates, which were incubated at 28°C for 48 h to let visible colonies develop. Thousands of colonies were pooled and grown in LBNS Km10 for two population doublings. Glycerol was then added to a final volume of 15% v/v, multiple 1-mL of library stocks were prepared and frozen at -80°C, and cell pellets were collected to extract genomic DNA for TnSeq mapping. Genomic locations of transposon insertions were mapped and linked to DNA barcodes using a variation of a previously described TnSeq protocol [36, 37]. To map transposon insertions and identify genes that are likely to be essential for viability (or nearly so) in LBNS, a previously established pipeline was used that applies previously published heuristics [38, 39] to distinguish likely-essential genes from genes that are too short or that are too repetitive to map insertions in. Briefly, for each protein-coding gene, the total read density in TnSeq (reads/nucleotides across the entire gene) and the density of insertion sites within the central 10–90% of each gene (sites/nucleotides) was computed. Genes that were very similar to other regions of the genome, or <275 bp, were excluded. For the remaining genes, read density was normalized by GC content by dividing each gene's read density by the running median of read density over a window of 201 genes (sorted by GC content). The insertion density was normalized so that the median gene's value was 1. Protein-coding genes were considered likely-essential if both the normalized insertion density (dens) and the normalized read density (normreads) were under 0.2 [39]. Analysis scripts are available at <https://bitbucket.org/berkeleylab/feba/>.

Genome sequencing and comparative analyses

Genomic DNA of the *P. sabiae* $\Delta 2$ strain was extracted using the GenElute Bacterial Genomic DNA kit (Sigma-Aldrich, MO, USA) and sequenced with MiSeq i100 (Illumina, Inc, CA, USA). Reads were mapped against the *P. sabiae* LMG23245^T (NCBI, Genbank accessions CP125295 to CP125298) using the CLC Genomic Workbench v.11.0 (QIAGEN, Aarhus, DEN). Functional classifications was done

using eggNOG-Mapper v2 5.0 [40]. The *repA* sequence (QEN71_RS30105) was searched (BLASTN) on the NCBI server to the “nr” (non-redundant) database (June 2025). Replicons carrying a homolog with 100% coverage and more than 80% identity were aligned using Proksee [41]. For building a phylogenetic tree, *P. sabiae* RepA homologs with a coverage of 100% and an identity >92% (BLASTP) were downloaded from NCBI and compared with the Molecular Evolutionary Genetics Analysis software 11 (MEGA11) [42]. Sequence alignment was performed using ClustalW [43] and phylogenetic inference was conducted using the maximum-likelihood method.

Phenotypic characterisation of *P. sabiae* $\Delta 2$

Aerobic growth of *P. sabiae* wild-type and $\Delta 2$ was assessed in 250 mL Erlenmeyer flasks containing 50 mL medium at 28°C with constant shaking at 180 rpm for 26 h. Colony forming units (CFU) were determined on LBNS and 0.5X LBNS plates. Microoxic growth was tested on modified ABS plates (A component diluted 1:9 with water) using GasPak EZ containers (BD; Becton, Dickinson and Company, Sparks, NV, USA) in which the oxygen level was maintained (6-16% O₂) with GasPak microaerophilic sachets that were replaced every 24 h. Anaerobic survival was tested as described previously [44] in ABS media. *In vitro* swimming motility assays were performed as previously described [45] and the swimming zone diameter was measured using ImageJ [46]. Competition assays *in vitro* were performed as previously described [31]. For intraspecies competition assay a ratio of 1:1 was used. The outcome of competition assays was measured both by imaging fluorescence with an Odyssey Fc Imager (LI-COR, USA) and by CFU determination.

Phenotypic microarrays

BIOLOG phenotype microarrays (PM1 and PM2; BIOLOG Inc, Hayward, CA, USA) were incubated at 28°C for 5 days. OD₆₀₀ of each well was measured every 12 h using an infinite M200Pro plate reader (Tecan, Männerdorf, CH). Data were normalized to the OD₆₀₀ values measured at the 0 h timepoint for each well. A phenotype was considered different when a minimum of a 50% difference in OD₆₀₀ was observed after 120 h, provided that the lowest value in the wild-type strain was at least 0.25. Validation was done in 96-well microtiter plates with AB liquid media supplemented with the different carbon sources (15 mM potassium succinate, 10 mM citric acid, 10 mM D-Glucose, 10 mM L-leucine, 5mM of D-trehalose; Sigma-Aldrich, MO, USA).

Transmission electron microscopy

Transmission electron microscopy (TEM) was performed using a FEI Tecnai G2 Spirit microscope (Field Electron and Ion Company, NE, USA) at the Center for Microscopy and Image Analysis in the University of Zurich. Acceleration voltage was aligned to 120 kV and a side-mounted digital camera Gatan Orius 1000 (Ametek, PA, USA) was used. *P. sabiae* strains were grown in 3 mL ABS (OD₆₀₀ 0.5) and spotted on a formvar/carbon 300-mesh copper grid (Electron Microscopy Sciences, PE, USA)

and stained using 1% w/v uranyl acetate (Sigma-Aldrich, MA, USA). Cell size was measured manually in Fiji [47].

Characterization of symbiotic properties

Mimosa caesalpiniiifolia seeds (Sunshine-Seeds, Ahlen, Germany) were surface sterilized for 15 min in 96% sulfuric acid (H₂SO₄, Merk, Darmstadt, Germany). Germinated seedlings were transferred to autoclaved 120 mL amber yogurt jars filled with vermiculite and one tablespoon of sand supplemented with Jensen medium [48], and were inoculated with the bacterial strain (10⁴ CFU) [13]. Plants were maintained in the greenhouse (25°C/22°C during day/night, 60% humidity, and 16/8 h light cycle). Nitrogenase activity was measured after 28 days by acetylene reduction assay (ARA) [49, 50]. For bacteroid reisolation, two nodules were selected per biological replicate and crushed in 100 µL of 0.5X LBNS containing 25% v/v glycerol and plated on LBNS plates [51].

Statistical analysis

Data visualization and analysis was performed in RStudio v2024.04.1+748 with R v4.2.0 [52] using *ggplot2* [53, 54], *ggpubr* [55] and *tidyverse* [56] packages. For the statistical analyses of competition assay, growth, CFU, and symbiotic properties, a one-way ANOVA with Tukey's multiple comparison was performed using *stat_compare_means* from the *ggpubr* package. For swimming, anaerobic survival, cell length and nodules occupancy, a Student's *t*-test was performed.

Results

Isolation of a *Paraburkholderia sabiae* LMG24235 strain unable to kill phytopathogens that has lost the 2.3 Mbp megaplasmid

We previously showed that *P. sabiae* LMG24235^T efficiently inhibits *P. syringae* pv. tomato DC3000 growth in a T6SS-1-dependent manner [31]. During a standard competition experiment, we observed that *P. sabiae* had lost its ability to kill DC3000 (Pto) (Fig. 1, Supplementary Fig. 1). PCR with primers targeting the T6SS-1 gene cluster on chromosome 2 (Supplementary Table 1) confirmed the absence of the T6SS-1 cluster in this spontaneously obtained strain. However, genes present on chromosome 1, plasmid 1 and plasmid 2 (QEN71_RS43065) could be amplified (Supplementary Fig. 2), suggesting that the non-killing *P. sabiae* strain had lost its second replicon. Genome sequencing confirmed that the non-killing strain had lost chromosome 2 (Fig. 2A), which contains 23% of the genes in the *P. sabiae* genome and was therefore named *P. sabiae* $\Delta 2$.

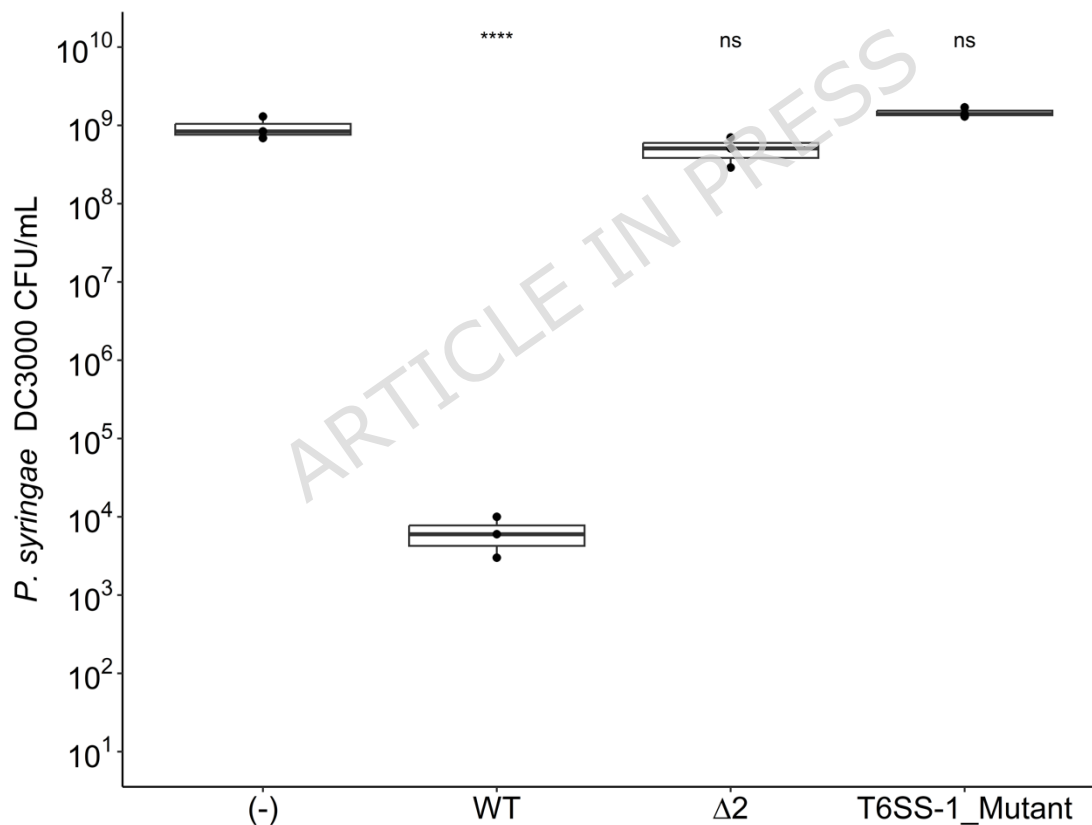
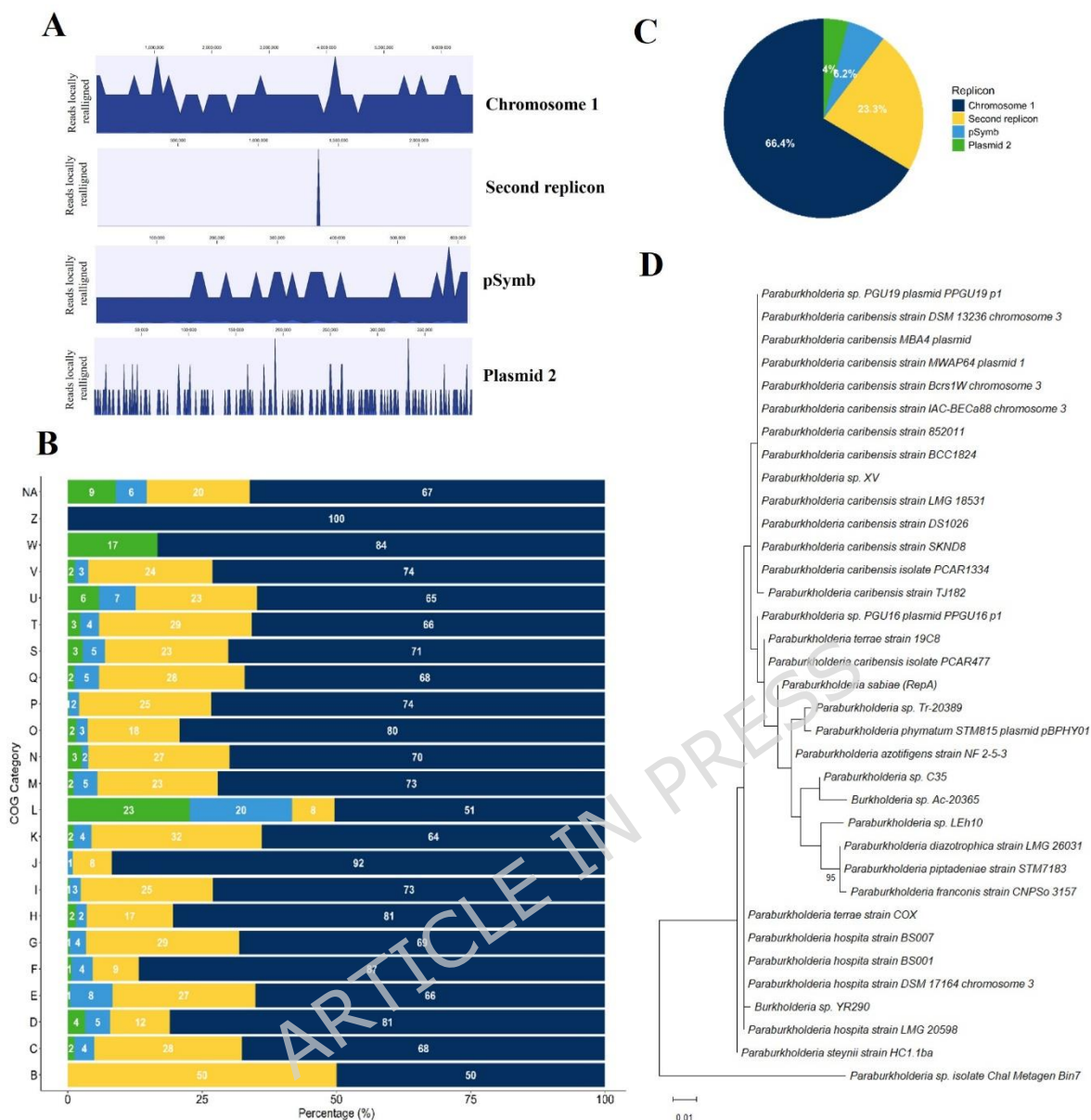


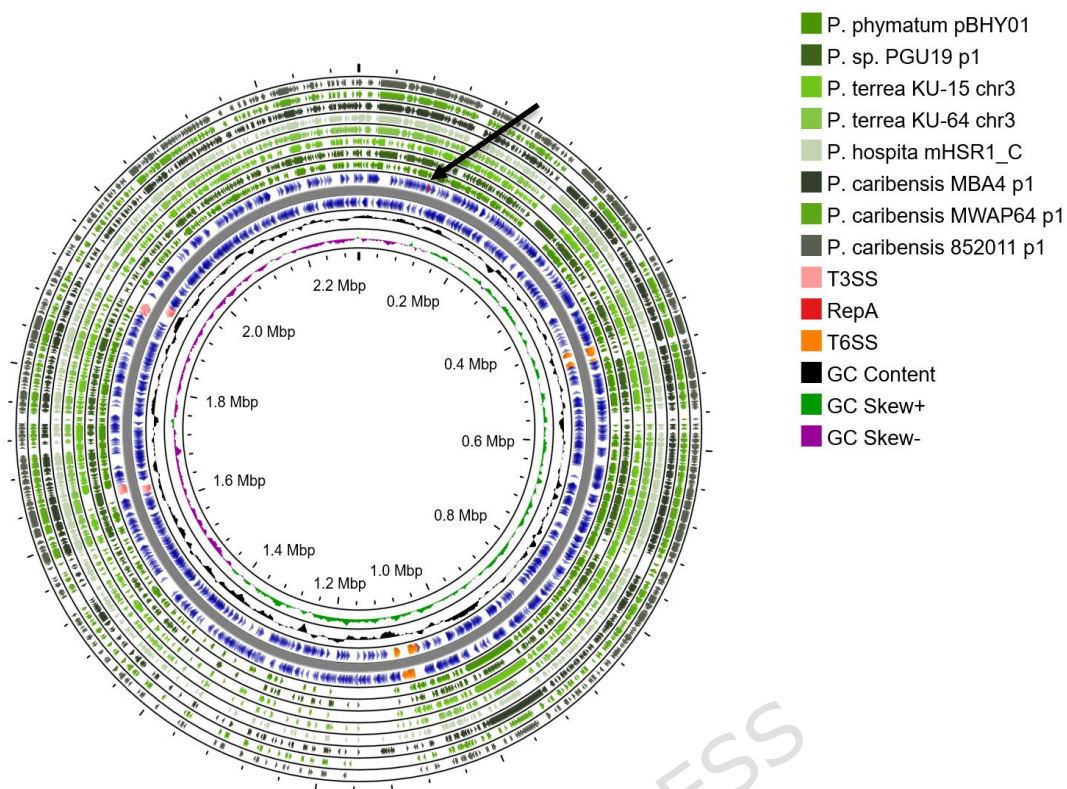
Fig. 1: Discovery of a non-killing *P. sabiae* strain. Competition assay performed with *P. sabiae* wild type (WT), the non-killing strain ($\Delta 2$), and a T6SS-1 mutant (T6SS-1) as attackers against *P. syringae* DC3000 as target. CFU/mL of the target strain *P. syringae* DC3000 is shown. Three biological replicates were performed. An ANOVA with Tukey's multiple-comparison test was performed to assess the statistical significance of differences compared to *P. syringae* incubated without attacker (-) (ns: not significant, $p \geq 0.05$; ****, $p \leq 0.0001$).

Given its dispensability, we propose to refer to the second replicon as a large megaplasmid rather than a chromosome. Moreover, as previously reported by our group, this replicon lacks rRNA genes [31]. To assess the general function encoded in the second replicon, the representation of each Cluster of Orthologous Group (COG) category on every *P. sabiae* replicon was analysed using the eggNOG database [40, 57]. Category K was over-represented on the second replicon compared to chromosome 1 (Fig. 2B, Supplementary Fig. 3). Moreover, genes related to carbohydrate transport and metabolism (category G), signal transduction mechanisms (T), amino acid transport and metabolism (E), energy production and conversion (C), lipid transport and metabolism (I), inorganic ion transport and metabolism (P), motility (N) and secondary metabolites biosynthesis, transport and catabolism (Q) were over-represented on the second replicon (Supplementary Fig. 3). In contrast, genes involved in translation, ribosomal structure and biogenesis (J) and nucleotide transport and metabolism (F) were under-represented on the second replicon (Fig. 2B, Supplementary Fig. 3). The number of genes belonging to the replication, recombination and repair category (L) was also low on the second replicon (8%) compared to the rest of the genome (Fig. 2C).

Bioinformatic analysis of the replication and partitioning system of the second replicon revealed the presence of *repA* (QEN71_RS30105), *parA* (QEN71_RS30110) and *parB* (QEN71_RS30115). Together, these genes enable the regulated partitioning of multipartite genomes such as those found in *Burkholderia* species [58, 59]. RepA (QEN71_RS30105) of the *P. sabiae* second replicon showed high similarity to RepB of plasmid pBPHY01 from *P. phymatum* STM815^T (PHY_RS35745, 100% coverage and 87.13% identity, BLASTN) and the phylogenetic analysis suggests a shared evolutionary origin (Fig. 2D). Alignment of replicons of closely related *Paraburkholderia* strains carrying highly similar *repA* genes showed strong conservation of genes adjacent to the replication origin, whereas genes located on the opposite side of the replication origin showed greater variability (Fig. 2E). Consistently, the second replicon of *P. sabiae* displayed high similarity to megaplasms identified in other *Paraburkholderia* strains (Fig. 2E, Supplementary Fig.4).



E



232

Fig. 2: The second replicon is absent in the non-killing *P. sabiae* strain ($\Delta 2$) **A:** Mapping of DNA sequencing from non-killing *P. sabiae* ($\Delta 2$). **B:** Distribution of COG categories across the *P. sabiae* genome using eggNOG-Mapper v2 5.0. Different colors represent the different replicons. Green is the plasmid 2, turquoise is pSymb, yellow is second replicon (megaplasmid), and dark blue is chromosome 1. The different categories are indicated by the following letters: B, chromatin structure and dynamics; C, energy production and conversion; D, cell cycle control, cell division, chromosome partitioning; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; J, translation, ribosomal structure and biogenesis; K, transcription; L, replication, recombination and repair; M, cell wall/membrane/envelope biogenesis; N, cell motility; O, posttranslational modification, protein turnover, chaperone; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport and catabolism; R, general function prediction only; S, function unknown; T, signal transduction mechanisms; U, intracellular trafficking, secretion, and vesicular transport; V, defence mechanisms; W, extracellular structures; Z, cytoskeleton; NA, not categorized. **C:** Size of the four replicons displayed in % compared to the whole genome size (9.9 Mb total genome, chromosome 1: 6.5 Mb, second replicon (megaplasmid): 2.3 Mbp, pSym: 6.15 Kbp, plasmid 2: 3.99 Kbp). **D:** Maximum likelihood phylogenetic tree based on RepA amino acid sequence. **E:** The closest replicons containing homologs of *repA* on *P. sabiae* LMG24235^T second replicon were aligned using Proksee. *P. sabiae* wild type second replicon is used as a backbone (dark blue). Different green shades indicate the different replicon of each species. The GC content is shown in black. *P. sabiae* LMG23245^T *repA* (QEN71_RS30105) is shown in red, indicated by the black arrow.

254

255

Randomly barcoded transposon mutagenesis sequencing (RB-TnSeq) supports the non-essentiality of the large megaplasmid

To identify essential genes in *P. sabiae* LMG24235^T, we constructed a RB-TnSeq mutant library (*Psabiae* _ML3) with 470,292 uniquely barcoded strains at 159,151 locations distributed across the genome (Fig. 3A). Of the 8,591 predicted protein-coding genes in *P. sabiae* LMG24235^T, we generated at least one centrally-located mutation in 7,764 genes (90.3% of the genome). For the remaining 827 genes we do not have gene data, which can be due to reasons such as gene size, gene duplicates, GC content, low sequencing coverage, or chance [39]. To identify which of those 827 genes are likely essential for viability (or nearly so) in the conditions in which we constructed the library (i.e. LBNS), we used a previously established pipeline [39]. Genes with redundancies or duplicates (dupScore>0) were ignored. Genes that are protein-coding, with dupScore=0 and ≥ 275 bp were classified as likely-essential or non-essential according to insertion density (Supplementary Table 2). In total, 336, 35, 7 and 1 genes were classified as likely essential (ess=TRUE, Supplementary Table 2) in chromosome 1, second replicon, pSymb and plasmid 2. This accounts for 5.79, 1.75, 1.48 and 0.3% of genes encoded on each replicon, respectively (Fig. 3B, Fig. 3C).

The 35 genes identified as likely essential on the second replicon (Fig. 3C, Supplementary Table 2) showed normalized insertion density (dens) and normalized read density (normreads) values below 0.2. This indicates that these genes are relatively small and/or have a low density of transposon insertion sites, potentially due to high GC content, as the *mariner* transposon inserts into TA sites. Examples among these 35 genes include the *parA* and *parB* (QEN71_RS30110-QEN71_RS30115), Type III secretion system (T3SS) related genes (QEN71_RS36795, QEN71_RS36815, QEN71_RS36825) and LysR-type regulators (QEN71_RS38935, QEN71_RS39730). Given that the second replicon is dispensable, these genes are not truly essential (see Discussion).

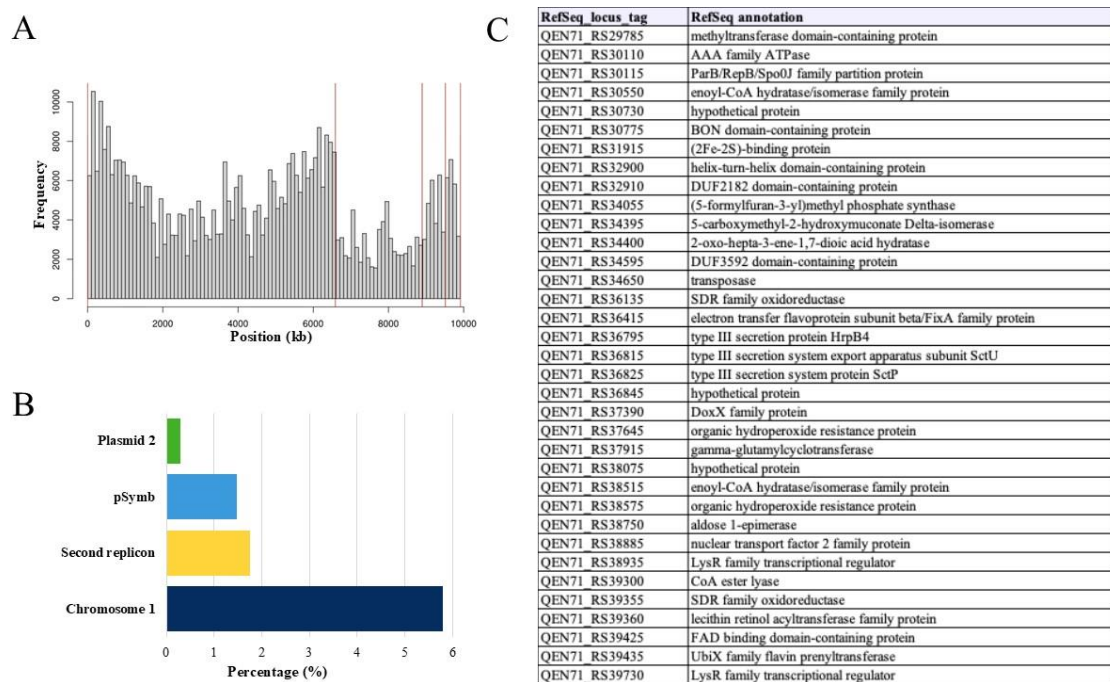


Fig. 3: Genome-wide mutant profiling by random barcoded transposon mutagenesis (RB Tn-Seq) in *P. sabiae* LMG24235^T reveals a lower percentage of essential genes on dispensable replicons. **A:** Histogram of transposon insertions in the *P. sabiae* _ML3 mutant library. Red lines indicate scaffold boundaries (replicons). **B:** Percentage of genes classified as likely essential on each replicon based on Tn-Seq. **C:** List of the 35 genes classified as likely essential on the second replicon. A complete list of genes is provided in Supplementary Table 2.

Absence of the second replicon leads to changes in carbon source utilization and to decreased survival in microoxic growth conditions

In complex and minimal media (ABS), *P. sabiae* $\Delta 2$ is not affected in growth (Supplementary Fig. 5). However, genes important for carbohydrate and amino acid metabolism and transport are over-represented on the second replicon (Fig. 2B). To assess the metabolic functions encoded by the second replicon, *P. sabiae* $\Delta 2$ and wild-type strain were grown on 190 different carbon sources (PM1 and PM2 plates) (Fig. 4A). *P. sabiae* $\Delta 2$ was impaired in growth in the presence of D-trehalose, citric acid and L-leucine as a sole carbon sources. To confirm these results, *P. sabiae* $\Delta 2$ was grown with AB minimal medium supplemented with 10 mM citric acid, 10 mM L-leucine, or 5 mM of D-trehalose. Glucose (10 mM) was used as negative control (Fig. 4B). While *P. sabiae* $\Delta 2$ could not use L-leucine as a carbon source and was impaired in growth on D-trehalose as compared to *P. sabiae* wild type (Fig. 4B), the growth defect of the $\Delta 2$ strain in citric acid was just delayed. As previously mentioned, several genes important for energy production and conservation (C) are over-represented on the second replicon. A search for terminal cytochrome oxidases in the *P. sabiae* genome revealed two gene clusters on the second replicon potentially encoding terminal oxidases (QEN71_RS37515-QEN71_RS37520 and QEN71_RS36885-QEN71_RS36900). These clusters resemble gene clusters in *Burkholderia*

cenocepacia H111: QEN71_RS37515-QEN71_RS37520 is similar to the cyanide-insensitive *bd*-type terminal oxidase Cio-2 (plasmid 3) and QEN71_RS36885-QEN71_RS36900 to the heme-copper cytochrome *bo*₃ ubiquinol oxidase [60]. Accordingly, microoxic growth of *P. sabiae* $\Delta 2$ on minimal medium plates containing succinate as carbon source was much slower compared to the wild type (Fig. 4C: 3-4). After 4 days of incubation in a microoxic jar, only *P. sabiae* wild-type colonies were visible. *P. sabiae* $\Delta 2$ colonies appeared only after 6 days of incubation. This growth delay was not observed on minimal medium plates under aerobic conditions (Fig. 4C: 1-2). We also looked at the ability of the obligate aerobe *P. sabiae* wild-type and $\Delta 2$ to survive under anaerobic conditions in minimal medium and observed that wild-type cells had a survival rate of 30%, while the survival rate of *P. sabiae* $\Delta 2$ dropped to 10% (Fig. 4D). These results demonstrate that the second replicon provides *P. sabiae* with a growth advantage in minimal media and in environments with low or no oxygen availability.

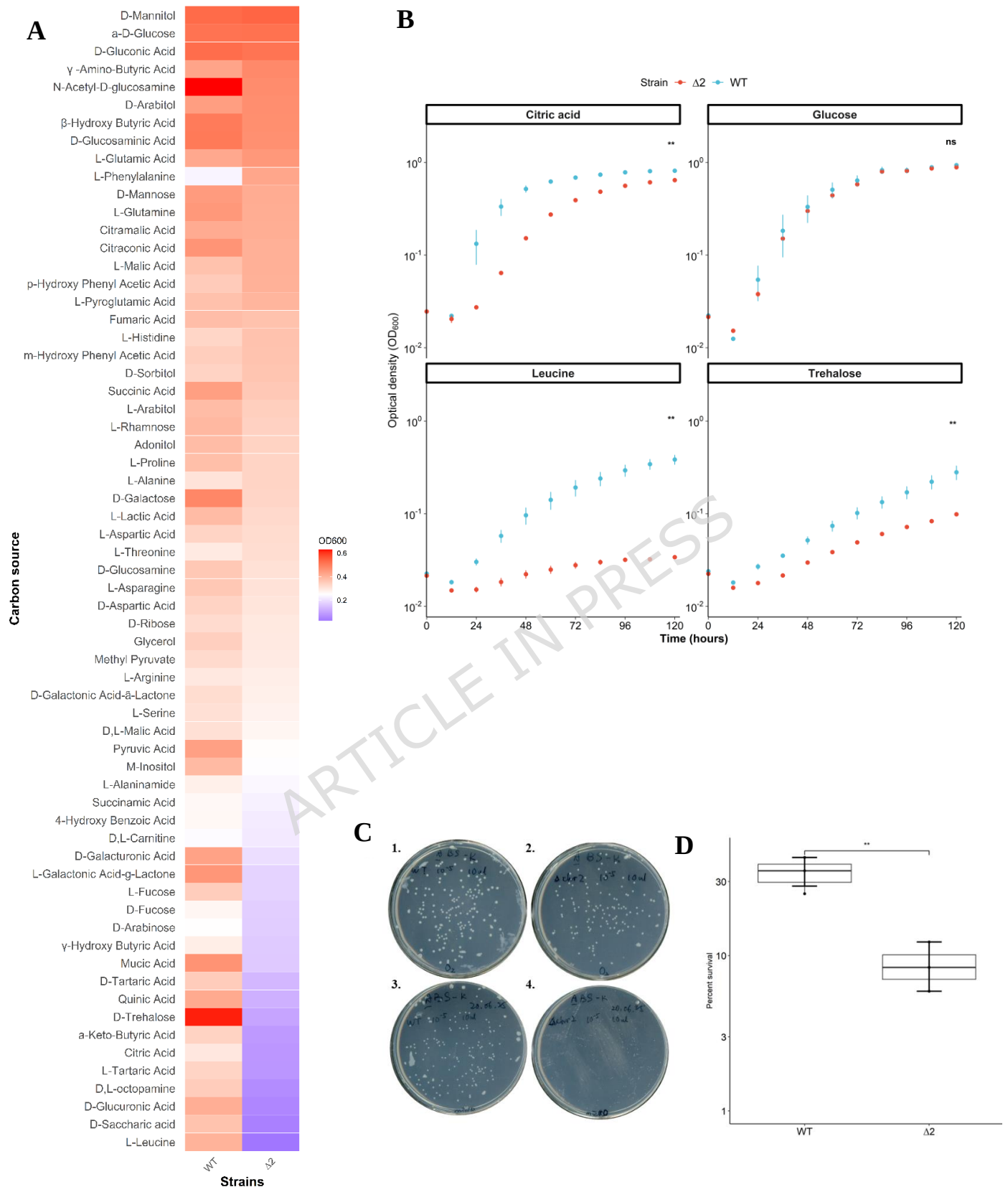


Fig. 4: Carbon source utilization of *P. sabiae* wild type (WT) and $\Delta 2$ **A:** Heatmap from the growth of *P. sabiae* WT and $\Delta 2$ in BIOLOG Phenotype Microarrays with different carbon sources (PM1-2). **B:** Growth of *P. sabiae* WT (blue) and $\Delta 2$ (red) in AB minimal media with different carbon sources (citric acid, glucose, leucine, trehalose) (N=3, n=2, Wilcox test, ns: not significant, p -value > 0.05, ** p -

value ≤ 0.01). **C:** Growth of *P. sabiae* wild type (1 and 3), and $\Delta 2$ (2 and 4) on ABS plates in aerobic (1 and 2) or microoxic conditions (2 and 4) for 4 days. **D:** Percentage survival of *P. sabiae* wild type (WT) and $\Delta 2$ under anoxic growth conditions in ABS media for 5 days (N= 3, n=2, *t*-test, ****p*-value ≤ 0.01).

***P. sabiae* $\Delta 2$ is impaired in swimming**

Since 27% of the *P. sabiae* genes involved in motility (Category N) are present on the second replicon, the ability of *P. sabiae* $\Delta 2$ to move was assessed and compared to the wild-type strain. In complex medium, swimming ability of *P. sabiae* $\Delta 2$ was reduced by 40% compared to the wild type (Fig. 5). To determine if this effect was due to impacts on cell morphology, the wild type and the *P. sabiae* $\Delta 2$ mutant were examined by TEM. Both strains possess one or more flagella (Supplementary Fig. 6A-B). However, the cell length of *P. sabiae* $\Delta 2$ was significantly shorter than that of the wild type (Supplementary Fig. 6C). Bioinformatic analysis revealed that 16 genes related to flagellar synthesis and regulation are present on the second replicon. In comparison, chromosome 1 contains three times as many flagellar genes, including the main components (*fliE*, *fliF*, *fliGH IJKLMN O P Q R*), and *fliA*, a flagellum-specific sigma factor. All flagellar genes present on the second replicon are also found on chromosome 1.

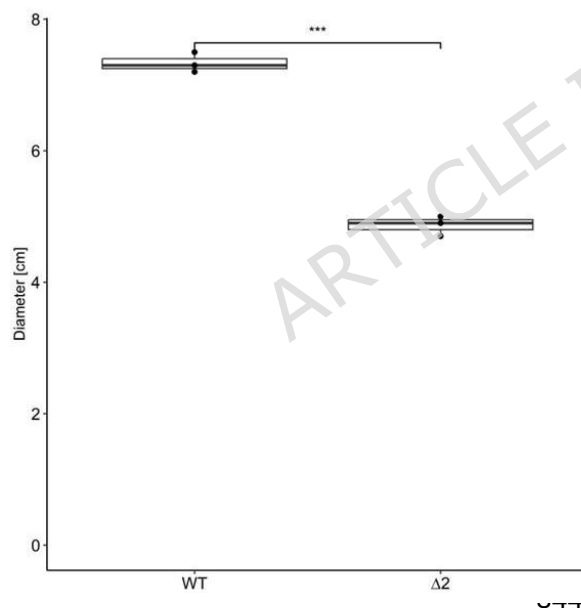
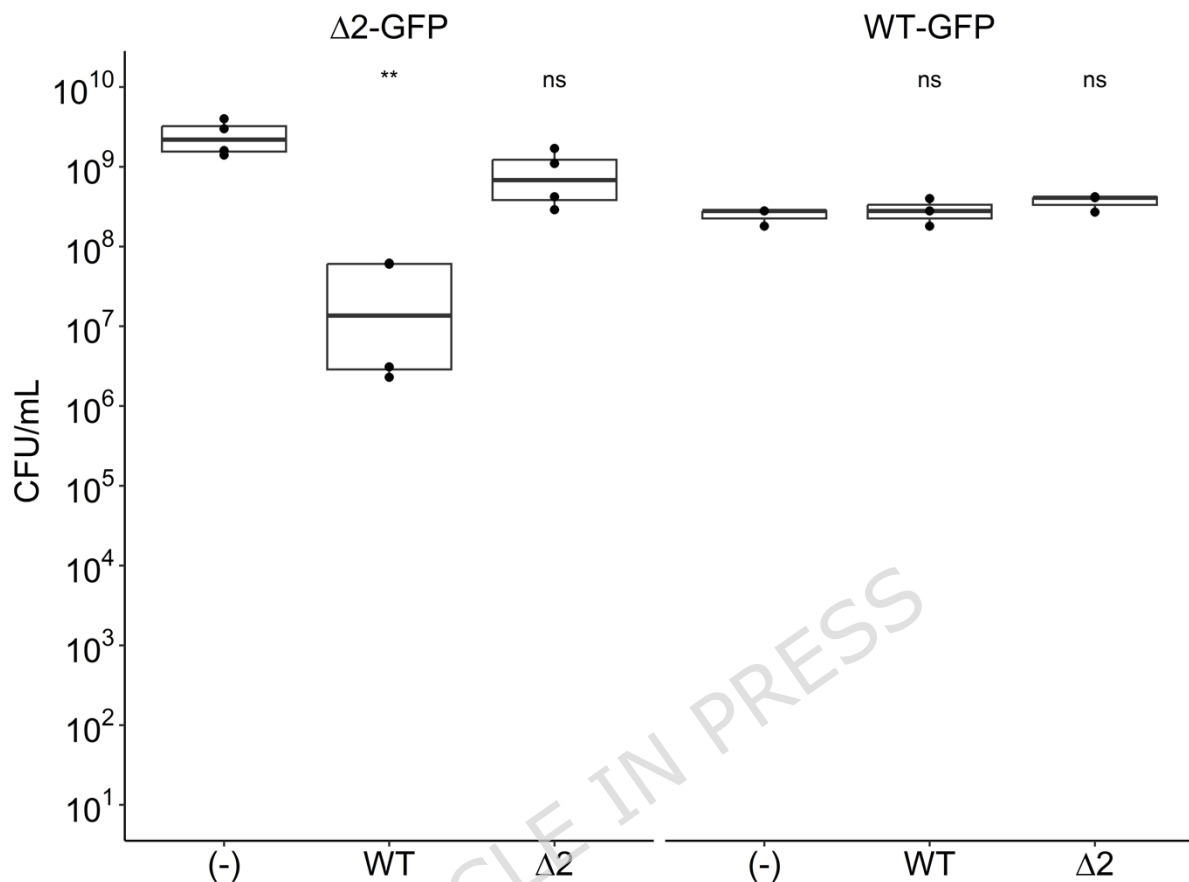


Fig. 5: Swimming motility of *P. sabiae* wild type (WT) and $\Delta 2$ in 0.1X LBNS. The diameter of each halo was measured after 48h incubation at 28°C (N= 3, n=2, *t*-test ****p*-value ≤ 0.0001).

***P. sabiae* wild type outcompetes $\Delta 2$ in vitro**

The second replicon is enriched in secretion systems, including two T6SS and two T3SS. These secretion systems are large multiprotein molecular complexes, and producing and maintaining them carries a high energy burden for the cell. To evaluate if the $\Delta 2$ strain was affected in intraspecific competition, we tested a GFP-tagged *P. sabiae* wild type (WT-GFP) against a *P. sabiae* $\Delta 2$ mutant, and *P. sabiae* wild-type against a GFP-tagged $\Delta 2$ ($\Delta 2$ -GFP). *P. sabiae* wild type outcompeted the $\Delta 2$ strain

by 2 orders of magnitude (Fig. 6). The presence of GFP did not affect competitive ability. These results show that the absence of the second replicon reduces the competitiveness of the *P. sabiae* $\Delta 2$ strain.



***P. sabiae* $\Delta 2$ enters symbiosis with *Mimosa caesalpiniiifolia* but shows reduced nitrogenase activity**

To evaluate if the absence of the 2.3 Mbp large megaplasmid affects symbiosis, the wild-type and the *P. sabiae* $\Delta 2$ mutant strains were inoculated onto *M. caesalpiniiifolia*. Twenty-eight days post-inoculation, the number of nodules (Fig. 7A) and their dry weight (not shown) were similar between plants inoculated with wild-type and *P. sabiae* $\Delta 2$ strains. However, nitrogenase activity in nodules induced by *P. sabiae* $\Delta 2$ was reduced by 35% compared to nodules inoculated with wild-type strain (Fig. 7B). Nodule occupancy analyses further indicated that both strains colonized *M. caesalpiniiifolia* nodules at roughly equal rates (Supplementary Fig. 7).

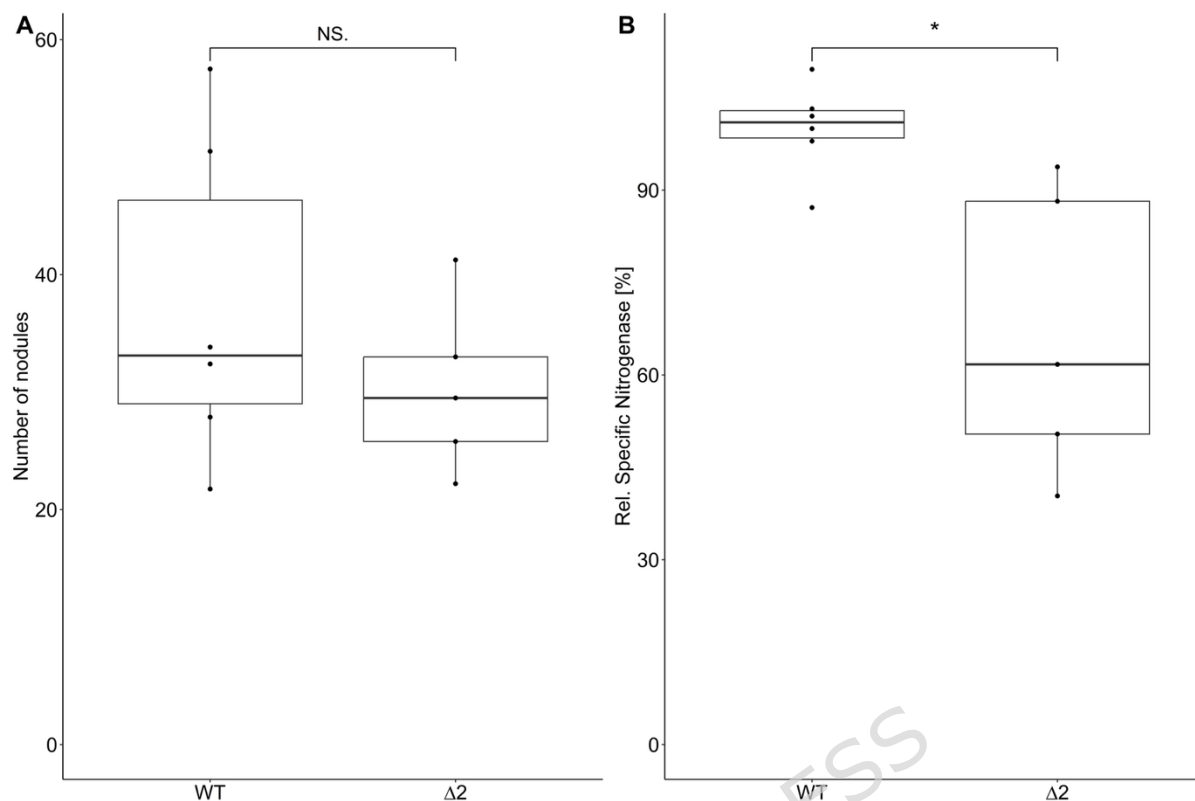


Fig. 7: Symbiotic properties of *M. caesalpiniiifolia* plants inoculated with *P. sabiae* WT and $\Delta 2$. **A:** Number of nodules per plant. **B:** Relative specific nitrogenase activity of *P. sabiae* WT (normalized as 100%) and $\Delta 2$. The statistical significance was calculated with one-way ANOVA with Tukey's test (NS: p-value > 0.05 *: p-value < 0.05).

Discussion

P. sabiae LMG24235^T possesses a large, multipartite genome, a feature often associated with highly versatile bacteria capable of adapting to diverse ecological niches. Such genome organization is thought to confer advantages for rapid adaptation to changing environments, particularly in bacteria that establish symbiotic relationships with eukaryotic hosts. *P. sabiae* was thought to possess two chromosomes (6.58 Mbp and 2.3 Mbp, respectively) and two plasmids, one of which is a symbiotic plasmid carrying genes required for plant symbiosis and nitrogen fixation. Several secretion systems were identified on the previously annotated second replicon (i.e. chromosome 2), among them the T6SS-1 shown by our group to mediate *P. sabiae*'s killing of phytopathogens [31]. During a competition assay using *P. sabiae* as an attacker, we isolated a mutant that had lost its killing ability. PCR and whole-genome sequencing revealed that this strain had lost its large second replicon of 2.3 Mbp, representing nearly one fourth of its genome. To our knowledge, this represents the largest spontaneous genome reduction reported in a free-living prokaryote. In rhizobia, spontaneous loss of the symbiotic plasmid (ranging from approximately 300 kb to 1.35 Mb) or substantial portions thereof, has been observed in the absence of selective pressure [62]. Similarly, within the closely related *Burkholderia cepacia* complex (Bcc), strains lacking the third replicon (typically 0.5-1.1 Mb in size)

have been reported [63]. In contrast to these spontaneous genome reduction events, engineered strains with defined large genomic deletions have been generated [62, 64–66].

Closer inspection of the genes encoded on the second replicon revealed the presence of *repA*, *parA* and *parB*, which are important for plasmid replication and partitioning. We show here that *P. sabiae* *repA* from the second replicon is highly similar to *repA/B* genes present on megaplasms in other *Paraburkholderia* strains, such as *P. phymatum* STM815^T (pBPHY01) [67], *P. caribensis* MBA4 and MWA61 (plasmid 1) [68] and *P. caribensis* DSM13236 (third replicon) [1, 69]. Moreover, similarly to the observation by Agnoli and colleagues [63], the replicons labelled as chromosome 3 in the tree presented in Figure 2D are most likely plasmids. In contrast to the consistent GC-skew observed on chromosome 1, the GC skew on the second replicon is weak [70, 71], and ribosomal RNA (rRNA) and aminoacyl-tRNA synthesis (tRNA) encoding genes, typically found on essential replicons, are absent. Construction and analysis of an RB-TnSeq library confirmed that the second replicon contains the fewest predicted essential genes (35 in total). However, because the second replicon is dispensable, these genes are not truly essential for viability. For instance, at least 17 genes had non-essential homologues (>33% homology) in other regions of the *P. sabiae* genome, and *parAB* genes were expected to lack transposon insertions, as disruption of these genes leads to loss of the replicon. Similarly, gene QEN71_RS34595, encoding a homolog of a T6SS immunity protein, would be expected to appear essential, as its deletion would result in self-intoxication. For some of the remaining genes, the reason for their apparent essentiality is unclear. However, it is important to note that the threshold of 0.2 used to classify genes as likely essential is arbitrary and lacks a direct biological basis. Although this criteria has been applied in more than 125 bacterial species to estimate genes that are likely to be essential (or nearly essential) for viability, it is a conservative threshold, meaning that there is a risk that genes are misidentified as likely-essential if 1) they are too short, 2) they have few TA sites (where the transposon lands in), or 3) they have low number of reads (due to sequencing) in the central part of the gene (10-90% of sequence). In conclusion, the TnSeq analysis strongly supports the hypothesis that the vast majority, if not all, of the genes required for growth of *P. sabiae* LMG24235^T are encoded on chromosome 1. To further confirm the non-essentiality of the second replicon, we examined whether homologues of the 35 genes predicted as essential were required for viability in other bacteria using the publicly available Fitness Browser website (<https://fit.genomics.lbl.gov>) [39], which compiles 9,262 genome-wide RB-TnSeq experiments across 57 bacteria and 2 archaea. For 33 of the 35 genes (Fig. 3C), we identified homologues that were not essential in closely related species, including *Paraburkholderia graminis* OAS925, *P. phytofirmans* PsJN, *P. bryophila* 376MFSha3.1 and *Pseudomonas* species [37, 39, 72]. We also compared the essential genes detected on each replicon of *P. sabiae* LMG23245^T with essential gene sets of *Burkholderia cenocepacia* J2315 and *Burkholderia thailandensis* E264 available in the Database of Essential Genes (DEG, <https://tubic.org/deg/public/index.php>) [73], which is based on TnSeq studies. None of the 35 genes

predicted as essential on the second replicon overlapped with conserved essential functions, such as replication, DNA repair, transcription, RNA translation and protein maturation [74]. Moreover, whereas *Burkholderia* genomes typically encode for two RpoD sigma factors, with the essential copy usually located on the second replicon [18, 75, 76], both *rpoD* genes in *P. sabiae* are localized on chromosome 1. Collectively, our findings are fully consistent with previous observations that chromosome 1 in *Burkholderia sensu lato* strains harbours house-keeping genes constituting the core genome, whereas additional replicons are often dispensable and primarily carry accessory functions [18, 63].

To evaluate the impact of the loss of the second replicon, we phenotypically characterized wild type and the $\Delta 2$ strain. We observed that the absence of the second replicon affected the utilization of D-trehalose and L-leucine as carbon sources. Genome inspection revealed the presence of *treF* (QEN71_RS30375) encoding a putative periplasmic alpha-alpha trehalase on the second replicon. Whether this gene itself is critical for trehalose utilization awaits investigation via generation of individual mutant strains. In contrast, genes coding for enzymes involved in L-leucine metabolic pathways were not found on the second replicon. These results could be exploited in the design of a growth medium containing L-leucine and/or D-trehalose as carbon source, selective for the presence of the second replicon.

COG analysis enumerated disparities in gene category representation between replicons that correlate to expected phenotypes in the $\Delta 2$ mutant. An enrichment of motility genes (category N) on the second replicon explains why $\Delta 2$ swims slower than WT (Fig. 5), while the loss of energy production and conversion genes (category C), also overrepresented on the second replicon, is demonstrated by *P. sabiae* $\Delta 2$'s impaired fitness under oxygen limitation (Fig. 4). More specifically, the latter deficit may be due to the loss of cytochromes encoded on second replicon: QEN71_RS37515-RS37520 encodes a potential cyanide-insensitive *bd*-type terminal oxidase (CioAB), while QEN71_RS36885-QEN71_RS36900 encodes a potential heme-copper cytochrome *bo*₃ ubiquinol oxidase (CyoABCD). Cytochrome *bd*-type oxidases generally have very high affinity for oxygen, enabling bacteria to maintain aerobic respiration under low-oxygen conditions. CioAB in *B. cenocepacia* H111 is also important for survival in the presence of cyanide-producing bacteria [44]. The presence of this oxidase in *P. sabiae* could give the strain an advantage in niches occupied by cyanide-producing soil bacteria, such as members of the *Pseudomonas* genus. Consistent with this, *P. sabiae* $\Delta 2$ was impaired under microoxic conditions, and survival assays under anaerobic conditions showed that *P. sabiae* $\Delta 2$ survived less well than the wild type (Fig. 4). Other *bd*-oxidases encoded on chromosome 1 (*cioABX*, QEN71_RS07815-QEN71_RS078309 and *cydABX*, QEN71_RS01605-QEN71_RS016215) likely contribute to residual growth under microoxic conditions, allowing survival within plant nodules.

All bacteria we re-isolated from nodules occupied by *P. sabiae* wild-type strain contained the second replicon, suggesting that the nodule is a selective environment and promotes the maintenance of the

second replicon. Future experiments, including dual RNA-sequencing of *M. caesalpinifolia* nodules colonized by *P. sabiae* may help identify: i) other genes on the second replicon required for efficient nitrogenase activity and ii) plant genes responding to the absence of the second replicon. The GFP tagged *P. sabiae* wild-type and $\Delta 2$ strains could be used to visualize nodule occupancy of both strains and to test if the second replicon provides a competitive advantage during nodulation. The fact that *in vitro* *P. sabiae* $\Delta 2$ is outcompeted by the wild-type strain suggests that the second replicon enhances *P. sabiae*'s competitiveness. Although this megaplasmid is non-essential, its persistence in the natural environment is likely maintained by strong selective pressure. Therefore, we predict that *P. sabiae* $\Delta 2$ would be less competitive in the soil against other bacteria due to the absence of several secretion systems (two T6SS and two uncharacterized T3SS), its reduced swimming motility and its reduced fitness in low oxygen environments. At the same time, the possibility of deleting the second replicon opens new perspectives on the potential advantages conferred by the multipartite genomes of *Paraburkholderia* species, particularly in this genus' adaptation to diverse environments, competition, colonization of new niches, and host interactions, which can ultimately explain their evolutionary success and ecological adaptability.

Acknowledgments

The authors acknowledge Dr. Sarah Paszti and Dr. Paula Bellés-Sancho for providing expert advice on cytochromes, COG and DE-Seq analyses, respectively. We acknowledge Dr. Maria Domenica Moccia from the Functional Genomics Center in Zurich (FGCZ) for assisting with sequencing of *P. sabiae*. Benedict Binder, Julie Ahrens, Sarah Jeggli, Giada Ornaghi, and Romina Steiger are acknowledged for their support in performing phenotypic tests and helping in plant test experiments. We thank Johanna Wahn for her assistance in using RStudio and Martinus Hjørungnes for his help in data visualization. The authors would also like to thank Morgan Price for mapping the randomly barcoded mutant library and processing data.

Funding

Marta Torres acknowledges funding from the European Union's Horizon 2020 research and innovation programme (Marie Skłodowska-Curie grant agreement No. 891542). Adam Deutschbauer was supported by m-CAFEs Microbial Community Analysis & Functional Evaluation in Soils (<https://mcafes.lbl.gov>), a Science Focus Area led by Lawrence Berkeley National Laboratory, supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231. This work was supported by the Swiss National Science Foundation (SNSF) with project number 310030_215282 to Gabriella Pessi.

Author contributions

The following author contributions were made.

Conceptualization: M.T. and G.P. Methodology: K.B., C.W., A.M.D., M.T. and G.P. Investigation: K.B., C.W., Y.L., D.G., L.E., S.J.H., M.T. and G.P. Data analysis: K.B., C.W., Y.L., D.G., A.M.D., L.E., S.J.H., M.T. and G.P. Visualization: K.B., C.W., Y.L., D.G., S.J.H. and M.T. Funding acquisition: M.T., A.M.D. and G.P. Writing – original draft: K.B. and G.P. Writing – review & editing: K.B., C.W., Y.L., D.G., A.M.D., L.E., S.J.H., M.T. and G.P.

Competing interests

The authors declare no competing interests.

Data availability

The datasets generated and analysed during this study are available in the NCBI short reads archive (SRA) repository, <https://www.ncbi.nlm.nih.gov/sra/PRJNA1423734>.

References

1. Sawana A, Adeolu M, Gupta RS. Molecular signatures and phylogenomic analysis of the genus *Burkholderia*: proposal for division of this genus into the emended genus *Burkholderia* containing pathogenic organisms and a new genus *Paraburkholderia* gen. nov. harboring environmental species. *Front Genet* 2014;5. <https://doi.org/10.3389/fgene.2014.00429>
2. Rojas-Rojas FU et al. The potential of *Paraburkholderia* species to enhance crop growth. *World J Microbiol Biotechnol* 2025;41:62. <https://doi.org/10.1007/s11274-025-04256-3>
3. Barrett CF, Parker MA. Prevalence of *Burkholderia* sp. nodule symbionts on four mimosoid legumes from Barro Colorado Island, Panama. *Systematic and Applied Microbiology* 2005;28:57–65. <https://doi.org/10.1016/j.syapm.2004.09.002>
4. Barrett CF, Parker MA. Coexistence of *Burkholderia*, *Cupriavidus*, and *Rhizobium* sp. Nodule Bacteria on two *Mimosa* spp. in Costa Rica. *Applied and Environmental Microbiology* 2006;72:1198–1206. <https://doi.org/10.1128/AEM.72.2.1198-1206.2006>
5. Bellés-Sancho P et al. Nitrogen-Fixing Symbiotic *Paraburkholderia* Species: Current Knowledge and Future Perspectives. *Nitrogen* 2023;4:135–158. <https://doi.org/10.3390/nitrogen4010010>
6. Bournaud C et al. *Paraburkholderia piptadeniae* sp. nov. and *Paraburkholderia ribeironis* sp. nov., two root-nodulating symbiotic species of *Piptadenia gonoacantha* in Brazil. *International Journal of Systematic and Evolutionary Microbiology* 2017;67:432–440. <https://doi.org/10.1099/ijsem.0.001648>
7. Chen W-M et al. *Burkholderia mimosarum* sp. nov., isolated from root nodules of *Mimosa* spp. from Taiwan and South America. *International Journal of Systematic and Evolutionary Microbiology* 2006;56:1847–1851. <https://doi.org/10.1099/ijms.0.64325-0>
8. Chen W-M et al. *Burkholderia sabiae* sp. nov., isolated from root nodules of *Mimosa caesalpiniiifolia*. *International Journal of Systematic and Evolutionary Microbiology* 2008;58:2174–2179. <https://doi.org/10.1099/ijms.0.65816-0>
9. Dall’Agnol RF et al. *Paraburkholderia nodosa* is the main N₂-fixing species trapped by promiscuous common bean (*Phaseolus vulgaris* L.) in the Brazilian ‘Cerradão’. *FEMS Microbiol Ecol* 2016;92:fiw108. <https://doi.org/10.1093/femsec/fiw108>

10. Elliott GN et al. *Burkholderia phymatum* is a highly effective nitrogen-fixing symbiont of *Mimosa* spp. and fixes nitrogen ex planta. *New Phytologist* 2007;**173**:168–180.
11. Elliott GN et al. *Burkholderia* spp. are the most competitive symbionts of *Mimosa*, particularly under N-limited conditions. *Environmental Microbiology* 2009;**11**:762–778. <https://doi.org/10.1111/j.1462-2920.2008.01799.x>
12. Gyaneshwar P et al. Legume-Nodulating Betaproteobacteria: Diversity, Host Range, and Future Prospects. *MPMI* 2011;**24**:1276–1288. <https://doi.org/10.1094/MPMI-06-11-0172>
13. Lardi M et al. Competition Experiments for Legume Infection Identify *Burkholderia phymatum* as a Highly Competitive β -Rhizobium. *Front Microbiol* 2017;**8**. <https://doi.org/10.3389/fmicb.2017.01527>
14. Paulitsch F et al. *Paraburkholderia quartelaensis* sp. nov., a nitrogen-fixing species isolated from nodules of *Mimosa gymnas* in an ecotone considered as a hotspot of biodiversity in Brazil. *Arch Microbiol* 2019;**201**:1435–1446. <https://doi.org/10.1007/s00203-019-01714-z>
15. Sheu S-Y et al. *Burkholderia diazotrophica* sp. nov., isolated from root nodules of *Mimosa* spp. *International Journal of Systematic and Evolutionary Microbiology* 2013;**63**:435–441. <https://doi.org/10.1099/ijs.0.039859-0>
16. Sheu S-Y et al. *Burkholderia dipogonis* sp. nov., isolated from root nodules of *Dipogon lignosus* in New Zealand and Western Australia. *International Journal of Systematic and Evolutionary Microbiology* 2015;**65**:4716–4723. <https://doi.org/10.1099/ijsem.0.000639>
17. Bellés-Sancho P et al. Metabolomics and Dual RNA-Sequencing on Root Nodules Revealed New Cellular Functions Controlled by *Paraburkholderia phymatum* NifA. *Metabolites* 2021;**11**:455. <https://doi.org/10.3390/metabo11070455>
18. diCenzo GC, Mengoni A, Perrin E. Chromids aid genome expansion and functional diversification in the family *Burkholderiaceae*. *Mol Biol Evol* 2019;**36**:562–574. <https://doi.org/10.1093/molbev/msy248>
19. diCenzo GC, Finan TM. The divided bacterial genome: structure, function, and evolution. *Microbiology and Molecular Biology Reviews* 2017;**81**:10.1128/mmbr.00019-17. <https://doi.org/10.1128/mmbr.00019-17>
20. Harrison PW et al. Introducing the bacterial ‘chromid’: not a chromosome, not a plasmid. *Trends in Microbiology* 2010;**18**:141–148. <https://doi.org/10.1016/j.tim.2009.12.010>
21. Chen H et al. Proteome analysis demonstrates complex replicon and luteolin interactions in pSyma-cured derivatives of *Sinorhizobium meliloti* strain 2011. *ELECTROPHORESIS* 2000;**21**:3833–3842. [https://doi.org/10.1002/1522-2683\(200011\)21:17%253C3833::AID-ELPS3833%253E3.0.CO;2-I](https://doi.org/10.1002/1522-2683(200011)21:17%253C3833::AID-ELPS3833%253E3.0.CO;2-I)
22. López-Guerrero MG et al. Rhizobial extrachromosomal replicon variability, stability and expression in natural niches. *Plasmid* 2012;**68**:149–158. <https://doi.org/10.1016/j.plasmid.2012.07.002>
23. Chen W-M et al. Proof that *Burkholderia* Strains Form Effective Symbioses with Legumes: a Study of Novel *Mimosa*-Nodulating Strains from South America. *Applied and Environmental Microbiology* 2005;**71**:7461–7471. <https://doi.org/10.1128/AEM.71.11.7461-7471.2005>
24. Gehlot HS et al. An invasive *Mimosa* in India does not adopt the symbionts of its native relatives. *Ann Bot* 2013;**112**:179–196. <https://doi.org/10.1093/aob/mct112>
25. Liu X et al. *Burkholderia* and *Cupriavidus* spp. are the preferred symbionts of *Mimosa* spp. in Southern China. *FEMS Microbiol Ecol* 2012;**80**:417–426. <https://doi.org/10.1111/j.1574-6941.2012.01310.x>
26. Mishra RPN et al. Genetic diversity of *Mimosa pudica* rhizobial symbionts in soils of French Guiana: investigating the origin and diversity of *Burkholderia phymatum* and other beta-rhizobia. *FEMS Microbiol Ecol* 2012;**79**:487–503. <https://doi.org/10.1111/j.1574-6941.2011.01235.x>
27. Howieson JG et al. Novel *Burkholderia* bacteria isolated from *Lebeckia ambigua* – A perennial suffrutescent legume of the fynbos. *Soil Biology and Biochemistry* 2013;**60**:55–64. <https://doi.org/10.1016/j.soilbio.2013.01.009>
28. Lemaire B et al. Identification, origin, and evolution of leaf nodulating symbionts of *Sericanthe* (Rubiaceae). *J Microbiol* 2011;**49**:935–941. <https://doi.org/10.1007/s12275-011-1163-5>

29. Lemaire B et al. Biogeographical patterns of legume-nodulating *Burkholderia* spp.: from African Fynbos to Continental scales. *Applied and Environmental Microbiology* 2016;**82**:5099–5115. <https://doi.org/10.1128/AEM.00591-16>
30. Melkonian R et al. The geographical patterns of symbiont diversity in the invasive legume *Mimosa pudica* can be explained by the competitiveness of its symbionts and by the host genotype. *Environmental Microbiology* 2014;**16**:2099–2111. <https://doi.org/10.1111/1462-2920.12286>
31. Hug S et al. *Paraburkholderia sabiae* uses one Type VI Secretion System (T6SS-1) as a powerful weapon against notorious plant pathogens. *Microbiology Spectrum* 2023;**11**:e01622-23. <https://doi.org/10.1128/spectrum.01622-23>
32. Clark DJ, Maaløe O. DNA replication and the division cycle in *Escherichia coli*. *Journal of Molecular Biology* 1967;**23**:99–112. [https://doi.org/10.1016/S0022-2836\(67\)80070-6](https://doi.org/10.1016/S0022-2836(67)80070-6)
33. Shastri S et al. An efficient system for the generation of marked genetic mutants in members of the genus *Burkholderia*. *Plasmid* 2017;**89**:49–56.
34. Herrero M, de Lorenzo V, Timmis KN. Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria. *J Bacteriol* 1990;**172**:6557–6567. <https://doi.org/10.1128/jb.172.11.6557-6567.1990>
35. Liu H et al. Magic Pools: Parallel Assessment of Transposon Delivery Vectors in Bacteria. *mSystems* 2018. <https://doi.org/10.1128/msystems.00143-17>
36. Rubin BE et al. Species- and site-specific genome editing in complex bacterial communities. *Nat Microbiol* 2022;**7**:34–47. <https://doi.org/10.1038/s41564-021-01014-7>
37. Wetmore KM et al. Rapid Quantification of Mutant Fitness in Diverse Bacteria by Sequencing Randomly Bar-Coded Transposons. *mBio* 2015;**6**:10.1128/mbio.00306-15. <https://doi.org/10.1128/mbio.00306-15>
38. Rubin BE et al. The essential gene set of a photosynthetic organism. *Proceedings of the National Academy of Sciences* 2015;**112**:E6634–E6643. <https://doi.org/10.1073/pnas.1519220112>
39. Price MN et al. Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* 2018;**557**:503–509. <https://doi.org/10.1038/s41586-018-0124-0>
40. Cantalapiedra CP et al. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* 2021;**38**:5825–5829. <https://doi.org/10.1093/molbev/msab293>
41. Grant JR et al. Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Res* 2023;**51**:W484–W492. <https://doi.org/10.1093/nar/gkad326>
42. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol* 2021;**38**:3022–3027. <https://doi.org/10.1093/molbev/msab120>
43. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;**22**:4673–4680. <https://doi.org/10.1093/nar/22.22.4673>
44. Paszti S et al. Identification of key factors for anoxic survival of *B. cenocepacia* H111. *Int J Mol Sci* 2022;**23**:4560. <https://doi.org/10.3390/ijms23094560>
45. Lardi M et al. *Paraburkholderia phymatum* STM815 σ^{54} controls utilization of dicarboxylates, motility, and T6SS-b expression. *Nitrogen* 2020;**1**:81–98. <https://doi.org/10.3390/nitrogen1020008>
46. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 2012;**9**:671–675. <https://doi.org/10.1038/nmeth.2089>
47. Schindelin J et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012;**9**:676–682. <https://doi.org/10.1038/nmeth.2019>
48. Hahn M, Hennecke H. Localized mutagenesis in *Rhizobium japonicum*. *Molec Gen Genet* 1984;**193**:46–52. <https://doi.org/10.1007/BF00327412>
49. de Campos SB et al. Mutations in two *Paraburkholderia phymatum* Type VI Secretion Systems cause reduced fitness in interbacterial competition. *Front Microbiol* 2017;**8**. <https://doi.org/10.3389/fmicb.2017.02473>

50. Lardi M et al. Transcriptome analysis of *Paraburkholderia phymatum* under nitrogen starvation and during symbiosis with *Phaseolus vulgaris*. *Genes* 2017;**8**:389. <https://doi.org/10.3390/genes8120389>
51. Bellés-Sancho P et al. *Paraburkholderia phymatum* Homocitrate Synthase NifV Plays a Key Role for Nitrogenase Activity during Symbiosis with Papilionoids and in Free-Living Growth Conditions. *Cells* 2021;**10**:952. <https://doi.org/10.3390/cells10040952>
52. Posit team. RStudio: Integrated Development Environment for R. Boston, MA: Posit Software, PBC, 2024.
53. Schloerke B et al. GGally: Extension to 'ggplot2'. 2024.
54. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
55. Kassambara A. ggpubr: 'ggplot2' Based Publication Ready Plots. 2023.
56. Wickham H et al. Welcome to the Tidyverse. *Journal of Open Source Software* 2019;**4**:1686. <https://doi.org/10.21105/joss.01686>
57. Huerta-Cepas J et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 2019;**47**:D309–D314. <https://doi.org/10.1093/nar/gky1085>
58. Drevinek P et al. Diversity of the *parB* and *repA* genes of the *Burkholderia cepacia* complex and their utility for rapid identification of *Burkholderia cenocepacia*. *BMC Microbiol* 2008;**8**:44. <https://doi.org/10.1186/1471-2180-8-44>
59. Dubarry N, Pasta F, Lane D. ParABS systems of the four replicons of *Burkholderia cenocepacia*: new chromosome centromeres confer partition specificity. *J Bacteriol* 2006;**188**:1489–1496. <https://doi.org/10.1128/JB.188.4.1489-1496.2006>
60. Paszti S et al. Insights into the diverse roles of the terminal oxidases in *Burkholderia cenocepacia* H111. *Sci Rep* 2025;**15**:2390. <https://doi.org/10.1038/s41598-025-86211-8>
61. Rosenberg C et al. Genes controlling early and late functions in symbiosis are located on a megaplasmid in *Rhizobium meliloti*. *Molec Gen Genet* 1981;**184**:326–333. <https://doi.org/10.1007/BF00272926>
62. Oresnik IJ et al. Megaplasmid pRme2011a of *Sinorhizobium meliloti* is not required for viability. *Journal of Bacteriology* 2000. <https://doi.org/10.1128/jb.182.12.3582-3586.2000>
63. Agnoli K et al. Exposing the third chromosome of *Burkholderia cepacia* complex strains as a virulence plasmid. *Mol Microbiol* 2012;**83**:362–378. <https://doi.org/10.1111/j.1365-2958.2011.07937.x>
64. Westers H et al. Genome engineering reveals large dispensable regions in *Bacillus subtilis*. *Mol Biol Evol* 2003;**20**:2076–2090. <https://doi.org/10.1093/molbev/msg219>
65. Pósfai G et al. Emergent properties of reduced-genome *Escherichia coli*. *Science* 2006;**312**:1044–1046. <https://doi.org/10.1126/science.1126439>
66. Kolisnychenko V et al. Engineering a reduced *Escherichia coli* genome. *Genome Res* 2002;**12**:640–647. <https://doi.org/10.1101/gr.217202>
67. Moulin L et al. Nodulation of legumes by members of the β -subclass of Proteobacteria. *Nature* 2001;**411**:948–950. <https://doi.org/10.1038/35082070>
68. Pan Y, Kong KF, Tsang JSH. Draft Genome Sequence of the Haloacid-Degrading *Burkholderia caribensis* Strain MBA4. *Genome Announcements* 2014;**2**:10.1128/genomea.00047-14. <https://doi.org/10.1128/genomea.00047-14>
69. Achouak W et al. *Burkholderia caribensis* sp. nov., an exopolysaccharide-producing bacterium isolated from vertisol microaggregates in Martinique. *Int J Syst Bacteriol* 1999;**49 Pt 2**:787–794. <https://doi.org/10.1099/00207713-49-2-787>
70. Arakawa K, Suzuki H, Tomita M. Quantitative analysis of replication-related mutation and selection pressures in bacterial chromosomes and plasmids using generalised GC skew index. *BMC Genomics* 2009;**10**:640. <https://doi.org/10.1186/1471-2164-10-640>
71. Breuer C. The GC Skew - Studying genome evolution. *Data Science Blog*. <https://nycdatascience.com/blog/student-works/using-the-gc-skew-to-analyze-bacterial-genome-evolution/> . .
72. Torres M et al. Bacterial fitness for plant colonization is influenced by plant growth substrate. *New Phytologist* 2025;**248**:3168–3190. <https://doi.org/10.1111/nph.70617>

73. Luo H et al. DEG 15, an update of the Database of Essential Genes that includes built-in analysis tools. *Nucleic Acids Res* 2021;**49**:D677–D686. <https://doi.org/10.1093/nar/gkaa917>
74. Gil R et al. Determination of the Core of a Minimal Bacterial Gene Set. *Microbiol Mol Biol Rev* 2004;**68**:518–537. <https://doi.org/10.1128/MMBR.68.3.518-537.2004>
75. Bellés-Sancho P et al. Tn-seq profiling reveals that NodS of the beta-rhizobium *Paraburkholderia phymatum* is detrimental for nodulating soybean. *Commun Biol* 2024;**7**:1706. <https://doi.org/10.1038/s42003-024-07385-x>
76. Higgins S et al. The Essential Genome of *Burkholderia cenocepacia* H111. *J Bacteriol* 2017;**199**:e00260-17. <https://doi.org/10.1128/JB.00260-17>

ARTICLE IN PRESS