1 **Diversity of isoprene-degrading bacteria in phyllosphere and soil communities from a**

2 **high isoprene-emitting environment: a Malaysian oil palm plantation**

3

4 Ornella Carrión[1‡*], Lisa Gibson[1‡], Dafydd M.O. Elias[2], Niall P. McNamara[2], Theo A. van Alen[3], Huub

5 J.M. Op den Camp[3], Christina Vimala Supramaniam[4], Terry J. McGenity[5], J. Colin Murrell[1*]

6 [1]School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4

7 7TJ, UK

8 [2]UK Centre of Ecology & Hydrology, Lancaster Environment Centre, Library Avenue, Bailrigg,

9 Lancaster, LA1 4AP, UK

10 [3]Department of Microbiology, Faculty of Science, IWWR, Radboud University Nijmegen,

11 Heyendaalseweg 135, NL-6525 AJ Nijmegen, The Netherlands

12 [4]School of Biosciences, Nottingham Centre of Sustainable Palm Oil, University of Nottingham-Malaysia,

13 Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia

14 [5]School of Life Sciences, University of Essex, Colchester, UK

15 [‡]These authors contributed equally to this work

16

17 *Corresponding authors:

18 J Colin Murrell, [1]School of Environmental Sciences, University of East Anglia

19 Norwich Research Park, NR4 7TJ, UK

20 E-mail: j.c.murrell@uea.ac.uk

21 Tel: (+44) 01603 592959

22

23 Ornella Carrión, [1]School of Environmental Sciences, University of East Anglia

24 Norwich Research Park, NR4 7TJ, UK

25 E-mail: o.carrion-fonseca@uea.ac.uk

26 Tel: (+44) 01603 592239

27

28

Abstract

Background: Isoprene is the most abundantly produced biogenic volatile organic compound (BVOC) on Earth, with annual global emissions almost equal to those of methane. Despite its importance in atmospheric chemistry and climate, little is known about the biological degradation of isoprene in the environment. The largest source of isoprene is terrestrial plants, and oil palms, the cultivation of which is expanding rapidly, are among the highest isoprene-producing trees.

Results: DNA stable isotope probing (DNA-SIP) to study the microbial isoprene-degrading community associated with oil palm trees revealed novel genera of isoprene-utilising bacteria including *Novosphingobium, Pelomonas, Rhodoblastus, Sphingomonas* and *Zoogloea* in both oil palm soils and on leaves. Amplicon sequencing of *isoA* genes, which encode the α-subunit of the isoprene monooxygenase (IsoMO), a key enzyme in isoprene metabolism, confirmed that oil palm trees harbour a novel diversity of *isoA* sequences. In addition, metagenome assembled genomes (MAGs) were reconstructed from oil palm soil and leaf metagenomes and putative isoprene degradation genes were identified. Analysis of unenriched metagenomes showed that *isoA*-containing bacteria are more abundant in soils than in the oil palm phyllosphere.

Conclusion: This study greatly expands the known diversity of bacteria that can metabolise isoprene and contributes to a better understanding of the biological degradation of this important but neglected climate-active gas.

Background

Isoprene (2-methyl-1, 3-butadiene), with atmospheric emissions of 400-600 $Tg \cdot y^{-1}$, is the most abundantly produced biogenic volatile compound (BVOC) on Earth. This is approximately one-third of the total volatile organic compounds (VOCs) emissions and almost equal to global annual emissions of methane [1, 2]. Due to its volatile nature and high reactivity, isoprene plays a complex role in atmospheric chemistry and hence, climate. In pristine environments with low levels of nitrogen oxides (NOx), isoprene reacts with hydroxyl radicals (OH) and reduces the oxidising capacity of the atmosphere, which in turn increases the lifetime of greenhouse gases such as methane. However, when NOx are present at high levels, as commonly found in urban areas, they react with isoprene to form nitrogen dioxide (NO_2) and increase the tropospheric levels of ozone, which has detrimental impacts on air quality, human health and impedes progress towards many of the UN sustainable development goals [3-6]. Conversely, the products of isoprene oxidation can form secondary aerosols and act as cloud condensation nuclei, resulting in a global cooling effect [7].

While there are some industrial sources of isoprene (0.8 $Tg \cdot y^{-1}$), primarily from the production of synthetic rubber [8], the vast majority of isoprene emissions (~90%) originate from terrestrial plants [9, 10], with small contributions from marine algae (0.1-12 $Tg \cdot y^{-1}$), bacteria, fungi and animals [9, 11-16]. The enzyme responsible for isoprene production in plants is isoprene synthase, the presence and activity of which can vary significantly even between closely related species [17-20]. In isoprene-emitting plants, isoprene is produced in the chloroplast via the methyl-erythritol 4-phosphate (MEP) pathway [21]. Isoprene synthase is responsible for converting dimethylallyl diphosphate (DMAPP) to isoprene. Despite the fact that up to 2% of the carbon fixed by isoprene-emitting plants contributes to the synthesis of isoprene [22-23], its role in

78   plants is not fully understood. It has been reported that isoprene improves the

79   resilience of plants to oxidative, thermal and biotic stresses [23-26]. However, the

80   molecular mechanisms behind these processes have not yet been fully elucidated. In

81   addition, it has been recently suggested that isoprene may play a role in regulating

82   gene expression in plants [27].

83   While the production and atmospheric fate of isoprene has been well studied,

84   biological consumption in the isoprene biogeochemical cycle remains relatively

85   unexplored. Field chamber and continuous-flow studies have shown that soils are a

86   biological sink for isoprene at environmentally relevant concentrations [28-30]. Several

87   bacterial strains capable of growing on isoprene as a sole carbon and energy source

88   have been isolated from soils, phyllosphere and aquatic environments (reviewed in

89   [31]). Most of these strains are Gram-positive Actinobacteria, although more recent

90   studies have led to the isolation of Gram-negative Proteobacteria expanding the

91   known diversity of isoprene-degrading bacteria [32]. All characterised isoprene-

92   utilising microorganisms contain six genes (*isoABCDEF*) that encode the isoprene

93   monooxygenase (IsoMO) enzyme, which catalyses the first step in the isoprene

94   degradation pathway. Adjacent genes *isoGHIJ* encode enzymes involved in the

95   subsequent steps of isoprene metabolism [33]. The IsoMO belongs to the soluble

96   diiron monooxygenase (SDIMO) family [34] and the α-subunit contains the diiron

97   centre at the putative active site. The gene encoding this IsoMO α-subunit, *isoA*, is

98   highly conserved in isoprene-utilising bacteria and is an excellent marker gene for

99   isoprene degraders [35, 36]. The development of probes targeting *isoA* has been a

100   successful approach to investigate the distribution, diversity and abundance of

101   isoprene degraders in several environments, including oil palm soils and leaves [35].

102   However, it is important to combine the use of *isoA* probes with other cultivation-

103   independent techniques such as DNA stable isotope probing (DNA-SIP) [37] to

104   investigate the diversity of active isoprene degraders in the environment and to better

105   assess the role that microbes play in the biogeochemical cycle of isoprene. Indeed,

106   previous DNA-SIP experiments with ^{13}C-labelled isoprene have led to the identification

107   of novel genera of isoprene degraders in phyllosphere and soil environments, such as

108   *Sphingopyxis, Ramlibacter* and *Variovorax* [32, 38]. In turn, the sequencing

109   information provided by these DNA-SIP experiments has allowed the design of

110   targeted cultivation strategies that resulted in the isolation of representative strains of

111   these novel genera of isoprene degraders [32], which now can be used as model

112   microorganisms to study how isoprene metabolism is regulated.

113   The oil palm tree (*Elaeis guineensis*) is one of the highest isoprene-producing trees,

114   with estimated emissions of 175 µg·g^{-1} (dry leaves)·h^{-1} [39]. Oil palm is a major crop

115   across South East Asia and is the source of 30% of the world's vegetable oil [40], and

116   in countries such as Malaysia it covers >85% of total agricultural land, with an ongoing

117   annual land usage increase of 6.9% attributed solely to oil palm cultivation [40, 41].

118   Therefore, the vast expansion of a single crop that emits such high amounts of

119   isoprene have raised serious concerns about the impact of oil palm plantations on air

120   quality [42]. Here, we combine cultivation-dependent techniques with DNA-SIP, *isoA*

121   and 16S rRNA gene amplicon sequencing, and focussed metagenomics, to study the

122   isoprene-degrading microbial communities associated with oil palm trees in a

123   Malaysian plantation, both from the phyllosphere and from the soil nearby.

124

125

126

Results and Discussion

Identification of active isoprene-degrading bacteria using DNA-SIP

Diversity of bacteria from soils in the vicinity of oil palm trees

Analysis of 16S rRNA gene amplicon sequences showed that the unenriched bacterial community from soils in the vicinity of oil palm trees was very similar across replicates, confirming that extraction and handling procedures were consistent (Fig S1 and Fig S2 show the relative abundance (RA) of 16S rRNA genes in these environmental samples). The unenriched soil microbial community (S T0) was mainly composed of Proteobacteria (40.8 ± 0.5% RA), Actinobacteria (13.1 ± 0.7%), Bacteroidetes (11.2 ± 1.4%) and Acidobacteria (10.8 ± 0.6%, Fig S1), all of which are dominant phyla in soils [43-46]. The most abundant genera were *Rhodoplanes* (5.9 ± 0.1%) and *Flavobacterium* (4.0 ± 0.9%; Fig S2).

Soils were then enriched with ^{13}C-isoprene to identify the active isoprene degraders in this environment through DNA-SIP (see Methods). Sequencing of the 16S rRNA genes of the ^{13}C-heavy fractions showed that, although there was considerable inter-sample variability, *Rhodoblastus* (10.2-33.7%) and *Pelomonas* (14.2-54.9%) were highly enriched in all soil replicates (S 13C H; Fig 1). *Novosphingobium* was one of the major genera labelled in the ^{13}C-heavy fractions of replicates 2 (S 13C H R2) and 3 (S 13C H R3) representing 47.8% and 24.5%, respectively. Finally, *Sphingomonas* dominated the isoprene-degrading community of replicate 3 (S 13C H R3) with a RA of 42.4% (Fig 1). These four genera had 19 to 90-fold higher RA in ^{13}C-heavy (S 13C H) than in the ^{13}C-light (S 13C L) soil fractions, and constituted 28.7-40.2% of the total microbial community of the unfractionated soils incubated with ^{13}C-isoprene (13C UF;

Fig S2), which strongly suggest that they are active isoprene degraders. As expected, *Novosphingobium, Pelomonas, Sphingomonas* and *Rhodoblastus* also dominated the ^{12}C-isoprene-incubated microbial community (S 12C L) and each genus had a very similar RA to those of the unfractionated ^{13}C-samples (S 13C UF; Fig S2).

Previous DNA-SIP experiments and cultivation-dependent studies have identified members of *Sphingomonadaceae* (*Sphingopyxis*) and *Comamonadaceae* (*Ramlibacter* and *Variovorax*) as isoprene degraders with a functional IsoMO [32, 36, 38]. However, this is the first evidence that other genera of these families such as *Sphingomonas, Novosphingobium* and *Pelomonas* are likely to be able to metabolise isoprene. In addition, *Rhodoblastus* is the first member of the *Beijerinckiaceae* family to be implicated in isoprene degradation. Therefore, it will be interesting to attempt to isolate representative strains of this genus in future studies in order to confirm this ability.

DNA-SIP and 16S rRNA gene amplicon sequencing showed that the oil palm soil harbours a distinct isoprene-degrading bacterial community from soils beneath temperate trees that emit high levels of isoprene, such as willow. DNA-SIP experiments using willow soil incubated with ^{13}C-labelled isoprene identified *Rhodococcus, Ramlibacter* and *Variovorax* as the major genera labelled in the ^{13}C-heavy fractions [32]. However, these genera represented <1% of the ^{13}C-heavy (S 13C H) fractions from oil palm soil. Also, other well-characterised isoprene-degrading microorganisms, such as *Gordonia, Nocardioides, Mycobacterium* and *Sphingopyxis* species [32, 47, 48], constituted only a small part of the isoprene-degrading community (<1%) from soils taken from the vicinity of oil palm trees.

175   Both unenriched samples and heavy DNA fractions from these soil incubations were

176   also subjected to metagenomic sequencing. Community composition of raw reads was

177   assessed with MetaPhlAn2 [49]. As MetaPhlAn2 uses a range of clade-specific marker

178   genes to assess the phylogeny of the metagenomics reads, results differed slightly

179   from those obtained using 16S rRNA gene amplicon sequencing analysis. According

180   to the phylogenetic analysis of the soil metagenomes, the unenriched soil (S T0)

181   community was dominated by Proteobacteria (81.5%), Actinobacteria (7.8%) and

182   Acidobacteria (6.3%; Fig S1), confirming the results obtained by the analysis of the

183   16S rRNA gene amplicon sequencing data. The most abundant bacteria in the

184   unenriched soils that could be classified at the genus level belonged to *Cupriavidus*

185   (14.2%), followed by *Pseudogulbenkiania* (13.2%) and *Burkholderia* (11.7%; Fig 1).

186   Metagenomic sequencing revealed that ^{13}C-heavy fractions from replicates 1 and 2 (S

187   13C H R1-2) were dominated by *Thiomonas* (34.5%) and *Gordonia* (18.2%), whereas

188   replicate 3 (S 13C H R3) had a higher abundance of *Gordonia* (47%) and *Sphingobium*

189   (21.8%; Fig 1). However, these genera represented <1% of the ^{13}C-heavy fractions in

190   the 16S rRNA gene amplicon sequencing data. It is not surprising to find members of

191   *Gordonia* dominating the isoprene-degrading community, since strains from this genus

192   have been shown to contain a complete isoprene degradation gene cluster [48].

193   However, this study provides the first evidence that *Thiomonas* and *Sphingobium*

194   species may be also able to catabolise isoprene.

195

196   *Diversity of bacteria from the phyllosphere of oil palm trees*

197   The bacterial community of unenriched oil palm leaf (L T0) samples was dominated

198   by Proteobacteria (74.5 ± 0.3%, Fig S1), which is not surprising since Proteobacteria

have been found to be the most abundant phylum in the phyllosphere of several plant species [50-53]. Firmicutes also constituted a major component of the unenriched bacterial community from oil palm leaves (22.1 ± 0.2%, Fig S1), as has been reported for some trees and agricultural plants [51, 52, 54-55]. The most abundant genera in the oil palm phyllosphere were *Acinetobacter* (26.4 ± 0.7%), followed by *Clostridium* (22.0 ± 0.2%) and *Enterobacter* (11.6 ± 0.2%; Fig 2).

The 16S rRNA gene amplicon sequencing data showed that the diversity of the isoprene-degrading community of the samples incubated with ^{13}C-isoprene (L 13C H) was highly consistent between replicates, with *Gordonia* (51.4 ± 9.4%) and *Zoogloea* (12.3 ± 2.2%) being the most abundant genera (Fig 2). The RA of *Gordonia* and *Zoogloea* was 84.9 and 58.2-fold higher in the ^{13}C-heavy (L 13C H) compared to the ^{13}C-light (L 13C L) fraction, respectively (Fig S3), indicating that they are active isoprene degraders. In addition, these two genera constituted 10.8% of the total microbial community of the ^{13}C-unfractionated (L 13C UF) samples and, as expected, were also highly abundant in the ^{12}C-isoprene-incubated (L 12C L) microbial community (13.6%; Fig S3).

Strains of *Gordonia* that grow on isoprene as sole carbon and energy source have been isolated previously from leaves of an oil palm tree in the Palm House of Kew Gardens, London [32]. However, although a number of SIP experiments with ^{13}C-isoprene have been performed with samples from a wide range of environments, including the phyllosphere, estuaries and soils, no members of the *Zoogloea* genus or the order *Rhodocyclales* have been identified as active isoprene degraders [32, 36, 38, 48]. Here, the identification of *Zoogloea* as an isoprene degrader indicates that the variety of microorganisms able to metabolise this important climate-active gas is greater than previously known.

224   *Rhizobium* also had a relatively high RA (8.5 ± 2.2%) in all replicates from ^{13}C-heavy

225   fractions (L 13C H) compared to the unenriched (L T0) samples (Fig 2). However, its

226   RA was 2.2-fold higher in the ^{13}C-heavy (L 13C H) than the ^{13}C-light (L 13C L)

227   fractions, which is the same ratio observed between the ^{12}C-heavy (L 12C H) and the

228   ^{12}C-light (L 12C L) fractions (Fig S3). Therefore, based on these data, and since no

229   strains of this genus have been isolated from this environment to corroborate its ability

230   to degrade isoprene, *Rhizobium* spp. cannot be yet confirmed as isoprene degraders.

231   Previous SIP experiments exploring the phyllosphere of other high isoprene-emitting

232   trees from temperate regions, such as poplar, identified *Rhodococcus* and *Variovorax*

233   as the major players in isoprene degradation [38]. However, in our experiment, the

234   phyllosphere from tropical oil palm trees yielded a distinct profile of active isoprene

235   degraders, with *Gordonia* and *Zoogloea* being the main genera enriched in the ^{13}C-

236   heavy fractions and *Rhodococcus* and *Variovorax* showing a low RA (2.1 ± 0.4% and

237   <1%, respectively). The RA of other well-characterised isoprene degraders such as

238   *Sphingopyxis, Ramlibacter, Nocardioides* or *Mycobacterium* [32, 48] also represented

239   <1% of the labelled bacterial community from oil palm leaves (L 13C H).

240   Phylogenetic analysis of the unenriched leaf (L T0) metagenomes confirmed that the

241   unenriched bacterial community of the oil palm phyllosphere was overwhelmingly

242   dominated by Proteobacteria (99.1% RA; Fig S1). At the genus level, metagenomics

243   analysis also supported the 16S rRNA gene amplicon sequencing data, since

244   *Acinetobacter* (40.8% RA) and *Enterobacter* (12.7% RA) were highly abundant in the

245   unenriched phyllosphere community, together with *Pantoea* (15.5% RA; Fig 2).

246   Metagenomic data showed that *Gordonia* constituted 93.7% of the isoprene-degrading

247   community of oil palm leaves (L 13C H R1-3; Fig 2), in accordance to the 16S rRNA

248 gene amplicon sequencing results. However, no *Zoogloea* sequences were identified

249 in the ^{13}C-heavy fractions in the metagenomic analysis probably due to the different

250 approach that MetaPhlAn2 uses to assign the phylogeny of the reads compared to the

251 16S rRNA gene amplicon analysis.

252

253 *Comparison of isoprene degraders from soils and phyllosphere of oil palm trees*

254 Studying the microbial diversity associated with plants is an essential step to

255 understand host-microbiome interactions. However, only a few studies comparing

256 microbial communities of phyllosphere and soils associated with the same plant

257 species have been conducted to date (e.g. [54, 56-57]). Here we show that the

258 unenriched bacterial communities from oil palm soils (S T0) and leaves (L T0) are

259 distinct even at the phylum level (Fig S1), as reported for other plant species [54],

260 although both soil and leaves are dominated by Proteobacteria. The 16S rRNA gene

261 amplicon sequencing data also revealed that the active isoprene-degrading bacteria

262 from soil samples were phylogenetically more diverse than those from the oil palm

263 phyllosphere (see above).

264 When comparing the unenriched soil (S T0) and leaf (L T0) communities, it is

265 interesting to note that although each major player in isoprene degradation was

266 present in these contrasting environments at similar RA (Table S1), they responded

267 differently to isoprene enrichment (S 13C H and L 13C H). This suggests that the

268 physiochemical conditions and/or interactions with other groups of microorganisms

269 shapes the composition of the isoprene-degrading community in a particular

270 environment.

271　In addition, unenriched soil and leaf oil palm metagenomes (S T0 and L T0) were

272　analysed for the presence and relative abundance of *isoA* genes. Metagenomic data

273　showed that *isoA*-containing bacteria were 5-fold more abundant in soil samples (1%

274　of Bacteria) than in the phyllosphere samples (0.2% of Bacteria). Metagenomes

275　obtained in previous studies of unenriched samples from high isoprene-emitting trees

276　from temperate regions, such as poplar [38] and willow [32], were also analysed for

277　comparison. Results showed that 0.7% of Bacteria from soil beneath a willow tree and

278　0.02% of Bacteria from poplar leaves contained *isoA* genes. These data, though

279　sparse, showed the same trend observed in the oil palm environment, with soils

280　containing greater numbers of bacteria with the genetic potential to degrade isoprene

281　than the phyllosphere. This finding is surprising considering the greater availability of

282　isoprene in the canopy than at ground-level [58] and indicates that soils could be a

283　more important sink for isoprene than previously thought.

284

285　<u>Recovery of metagenome assembled genomes (MAGs)</u>

286　Assembled contigs from soil and leaf metagenomes were used to reconstruct

287　metagenome assembled genomes (MAGs) using MaxBin2 [59]. A total of 20 MAGs

288　from soil and 52 from leaf samples were obtained (Table S2). From these, two MAGs

289　from soils and three from leaf ^{13}C-heavy DNA metagenomes with >75% completeness

290　and <10% contamination contained genes encoding homologous polypeptides to

291　IsoABCDEF (E<1e-40). MAGs containing IsoMO-encoding genes from soil

292　incubations were taxonomically classified as *Novosphingobium* and *Rhizobiales,* and

293　leaf MAGs were classified as *Gordonia*, *Zoogloeaceae* and *Ralstonia* (Table S3).

The *Novosphingobium* soil-associated MAG contained the full isoprene degradation gene cluster (*isoABCDEFGHIJ*) on a single contig along with *aldH1*, which encodes an aldehyde dehydrogenase [33]. However, no further accessory genes were recovered (Fig 3). The products of these genes shared 76.2-100% amino acid identity (Table S4) with the corresponding polypeptides from *Sphingopyxis* sp. OPL5, a *Sphingomonadales* strain isolated from oil palm [32]. When the diversity and abundance of *isoA* genes in the ^{13}C-heavy DNA fractions from soil samples were analysed by *isoA* amplicon sequencing, two amplicon sequence variants (ASVs), ASV44 and ASV11, were identified, closely related to the IsoA from the *Novosphingobium* MAG (>99% amino acid identity; Table S5). These two ASVs represented 7.9% of the *isoA* genes of the ^{13}C-heavy DNA fraction from replicate 2 and 11% from replicate 3, respectively (Fig 4).

A *Rhizobiales* MAG was also reconstructed from ^{13}C-heavy DNA soil samples. Although this MAG had a high completeness (97.6%) and low contamination (2.5%), it showed a high strain heterogeneity (79%; Table S3), indicating that the MAG originated from DNA from one or more closely related microorganisms, thus making identification at a higher resolution difficult. Despite this, a complete isoprene degradation gene cluster (*isoABCDEFGHIJ*) plus *aldH1* were located in a single contig (Fig 3). When these genes were translated, they exhibited an amino acid identity of 54.8-84.9% to the homologous proteins from *Sphingopyxis* sp. OPL5, except for *isoD*, the product of which was more closely related to IsoD from *Ramlibacter* sp. WS9 [32] (Table S4). *isoA* gene amplicon sequencing analysis revealed that the ASVs closely related to the IsoA from the *Rhizobiales* MAG (>71% amino acid identity; Table S5) dominated the *isoA*-containing bacterial community from all ^{13}C-heavy DNA soil replicates, comprising a total of 78 ± 6.8% of the *isoA* genes in these samples (Fig 4).

319 The *Gordonia* phyllosphere MAG was identified to the species level as *Gordonia*
320 *polyisoprenovorans* i37 [48] (average nucleotide identity; ANI: 96.7%) and contained
321 all IsoMO-encoding genes *isoABCDEF*, along with downstream genes *isoGHIJ* (Fig
322 3). This MAG also contained *aldH2, CoA-DSR* and *gshB*, which are accessory genes
323 often found within the isoprene degradation gene cluster in Gram-positive bacteria and
324 encode an aldehyde dehydrogenase, a CoA-disulfide reductase and a glutathione
325 synthetase, respectively [33]. All the genes associated with the isoprene degradation
326 pathway recovered in the *Gordonia* MAG encoded polypeptides that shared >94%
327 amino acid identity to the corresponding proteins from *Gordonia polyisoprenovorans*
328 i37, except for IsoB (82.1%) and AldH1 (83%; Table S4). When the ^{13}C-heavy DNA
329 fractions from leaf samples were analysed by *isoA* amplicon sequencing, several
330 ASVs with a high percentage of amino acid identity to the IsoA from the *Gordonia* MAG
331 (>93%; Table S5) were recovered. However, ASV1, which showed 100% amino acid
332 identity to IsoA from the *Gordonia* MAG, overwhelmingly dominated the *isoA*-
333 containing bacterial community, representing 91.9 ± 7.3% of the *isoA* genes in these
334 samples (Fig 4).

335 The remaining two phyllosphere MAGs, *Zoogloeaceae* and *Ralstonia*, contained
336 genes that encoded homologous polypeptides to IsoABCDEF, although they showed
337 a low amino acid identity to the corresponding IsoMO proteins from well-characterised
338 isoprene degraders (34.1-52.5%; Table S4). In addition, no gene homologues to
339 *isoGHIJ* were recovered from these MAGs (Fig 3). While the absence of homologues
340 to *isoGHIJ* and the low sequence identity to IsoABCDEF indicates that these bacteria
341 may harbour a novel isoprene degradation pathway, especially considering that these
342 MAGs were recovered from ^{13}C-heavy DNA fractions metagenomes, we cannot be
343 absolutely certain that they are from *bona fide* isoprene degraders. These *isoA*-like

344 sequences were not identified by the *isoA* amplicon analysis, suggesting that the

345 relatively high number of mismatches with the *isoA* primers prevented successful PCR

346 amplification. Therefore, further targeted isolations of these bacteria and/or expression

347 of these *isoABCDEF* genes in a heterologous host [38] are required to establish that

348 these are genuine isoprene degraders.

349 Finally, if *Zoogloeaceae* and *Ralstonia* microorganisms could be confirmed as *bona*

350 *fide* isoprene degraders, the abundance of bacteria with the potential to metabolise

351 isoprene would increase from 1% to 2% in oil palm soil and from 0.1% to 0.2% in

352 phyllosphere unenriched samples. Similarly, the RA of *isoA*-containing bacteria would

353 also increase in both willow soil (from 0.7% to 1.3% of Bacteria) and poplar leaf (from

354 0.02% to 0.05% of Bacteria) unenriched samples, indicating that isoprene degraders

355 could be more abundant in the environment than previously thought. However, more

356 samples from contrasting ecosystems need to be explored to support this hypothesis.

357

358 <u>Isolation and characterisation of *Variovorax* sp. OPL2.2</u>

359 Cultures set up using material from the DNA-SIP experiments from soils and leaves

360 were subcultured three times at 2-week intervals with 25 ppmv isoprene before plating

361 onto minimal medium with isoprene as sole carbon source (see Methods). A strain

362 belonging to the genus *Variovorax* isolated from leaf enrichments, was able to grow

363 on isoprene as sole carbon and energy source (Fig S4).

364 Although DNA-SIP experiments have shown that *Variovorax* plays an important role

365 in isoprene degradation in the phyllosphere [38], *Variovorax* sp. OPL2.2 is the first

366 strain of this genus isolated from a tropical environment. When genomic DNA of

367 OPL2.2 was screened for *isoA*, it yielded a PCR product which had a translated

368 sequence with 99.4% amino acid identity to IsoA from *Variovorax* sp. WS11, an

369 isoprene degrader isolated from willow soil [32].

370 The genome of this new isolate from oil palm, *Variovorax* sp. OPL2.2, was sequenced

371 using Illumina and Nanopore technologies to confirm that it contained a full isoprene

372 degradation gene cluster. Assembly of both Illumina and Nanopore reads and

373 downstream analysis with CheckM [60] revealed that the *Variovorax* sp. OPL2.2

374 genome was comprised of 50 contigs totalling 8.5 Mbp and had a 98.4%

375 completeness, 1.1% contamination and a GC content of 67.4%. Finally, after

376 automatic annotation by Prokka [61], 8,200 predicted coding sequences were found

377 in the *Variovorax* sp. OPL2.2 genome.

378 Genome analysis confirmed that *Variovorax* sp. OPL2.2 contained *isoABCDEF*

379 encoding IsoMO. *isoGHIJ* and *aldH1* genes, which are involved in the subsequent

380 steps of isoprene metabolism, were located upstream *isoABCDEF* in an identical

381 layout to those of many *bona fide* isoprene-degrading strains [38] (Fig 3). *garB*, which

382 encodes a glutathione disulfide reductase, was also located in the same gene cluster

383 (Fig 3). *isoABCDEFGHIJ, aldH1* and *garB*, encoded polypeptides with high amino acid

384 identity (99.7-100%) to those from *Variovorax* sp. WS11. Indeed, ANI analysis

385 (>99.9%) revealed that *Variovorax* sp. OPL2.2 is the same species as *Variovorax* sp.

386 WS11.

387

388 **Conclusions**

389 The area of agricultural land dedicated to cultivating oil palm has increased

390 dramatically over the last forty years and continues to increase annually due to palm

391 oil demand from the food industry, domestic products and biofuel production [42].

Growth of a high isoprene-emitting crop and the consequent reactions of isoprene with NOx have raised concerns about the impact of large oil palm plantations on air quality and the regional delivery of the UN sustainable development goals. Therefore, to better understand the role that microbes play in mitigating the effects of this climate-active gas, it is essential to examine the microbial diversity and abundance of isoprene-degrading bacteria. In this study, we show that oil palm harbours a unique and distinct community of isoprene degraders compared to other high isoprene-emitting trees from temperate environments and contains higher numbers of bacteria with the genetic potential to metabolise this climate-active gas, especially soils beneath oil palm trees. Isoprene-utilising bacteria from oil palm soils were phylogenetically more diverse than those from the oil palm phyllosphere, with *Novosphingobium, Pelomonas, Rhodoblastus*, *Sphingomonas* being found in soils and *Gordonia* and *Zoogloea* on oil palm leaves. Analysis of MAGs revealed that bins from these genera contained isoprene degradation gene clusters and amplicon sequencing data showed that oil palm trees contain novel *isoA* sequences, many of which were highly similar to *isoA* genes recovered from *Novosphingobium, Gordonia* or *Rhizobiales* MAGs. The discovery of novel isoprene-degrading bacteria enhances a more robust *isoA* database necessary to utilise the *isoA* probes and assess the distribution, abundance and activity of the isoprene degraders in the environment.

Methods

DNA-stable isotope probing

Three trees from different locations in an oil palm plantation in Palong (Negeri Sembilan, Malaysia) were sampled on 12th November 2018 (Table S6). Soil (50 g)

was collected at a depth of 0-5 cm within the palm circle of three 28-year old oil palm trees after removing vegetation from the surface of the soil. Five healthy leaflets from lower canopy fronds of the corresponding trees were sampled to allow comparison of the diversity of isoprene-degrading bacteria between the oil palm phyllosphere and adjacent soils. Soils and leaf samples were sent to the laboratory of Prof. Niall McNamara at the UK Centre for Ecology and Hydrology in Lancaster, where appropriate facilities to import soils into the UK were available, for processing. Cells were extracted from the soils as follows: 5 g of soils were resuspended in 50 ml sterile distilled water in a sterile 250 ml conical flask. Flasks were then shaken at 150 rpm for 30 min at room temperature to dislodge cells from soil particles. Soil suspensions were decanted into a 50 ml measuring cylinder and left undisturbed for 1 h to allow sedimentation of soil particles. The aqueous layer was decanted into a sterile flask. This treatment was repeated including a sonication step for 5 min in a water bath (Mettler Electronics) to optimise recovery of cells that were attached to soil particles. Soils particles were appropriately disposed of by autoclaving and the aqueous layers of both treatments were then combined and transported in sealed vials to University of East Anglia, where microcosms with soil washings were set up.

Soil DNA-SIP enrichments were set up in triplicate and consisted of 80 ml of combined soil washings in 2 litre air-tight bottles containing 25 ppmv of either ^{12}C or ^{13}C-labelled isoprene. Soil microcosms were incubated at 30 ˚C in the dark with agitation at 150 rpm. Five leaves per oil palm tree (the same trees used for the soil microcosms) were cut into approximately 10 cm length and 5 cm width and placed in sterile glass bottles containing 250 ml sterile distilled water. Leaf samples were sonicated for 5 min in a water bath (Mettler Electronics) and agitated at 150 rpm at room temperature for 1 h to dislodge microbial cells from plant material. Leaf washings were then filtered

through 0.22 μm cellulose nitrate membrane filters (Pall) to concentrate cells. Filters were washed with 40 ml of Ewers minimal medium [62]. Subsequently, filters were discarded, and the washings transferred to 2 litre air-tight bottles to form the basis of the leaf microcosms. Phyllosphere microcosms were then amended with 25 ppmv of either ^{12}C or ^{13}C-labelled isoprene. Leaf DNA-SIP enrichments were set up in triplicate and incubated at 30 ˚C in the dark with shaking (150 rpm). Consumption of isoprene by soil and leaf microcosms was monitored with a Fast Isoprene Sensor (Hills-Scientific) and replenished when the headspace concentration fell below 10 ppmv.

Sampling of soil and leaf DNA-SIP incubations was performed as follows: 10 ml aliquots were collected at T0 and after 5 days (12.5 μmol C assimilated $\cdot g^{-1}$) of incubation for soil microcosms. 10 ml aliquots from microcosms set up with leaf washings were collected at T0 and after 10 days (50 μmol C assimilated $\cdot g^{-1}$) of incubation. Soil and leaf aliquots were then spun down and supernatants were discarded. After that, pellets were resuspended with a 1 ml solution containing sodium phosphate and MT buffers included in the FastDNA Spin Kit for Soil (MP Biomedicals) and transferred to Lysing matrix E tubes to proceed with the DNA extraction according to the manufacturer's instructions. 0.5-2 μg DNA per sample was separated into heavy (^{13}C-labelled) and light (^{12}C-unlabelled) DNA by isopycnic ultracentrifugation as previously described [36]. DNA in each fraction was quantified using a Qubit dsDNA HS Assay kit (ThermoFisher Scientific) following the manufacturer's instructions. The density of each fraction was estimated by refractometry using a Reichert AR200 refractometer (Reichert Analytical Instruments). Heavy and light DNA fractions from each sample were identified by plotting DNA abundance vs fraction density (Fig S5) and used for subsequent downstream analysis.

466 <u>16S rRNA gene amplicon sequencing</u>

467 To investigate the bacterial diversity in samples from DNA-SIP experiments, 16S rRNA

468 genes of DNA extracted from unenriched (T0), unfractionated (UF), labelled (heavy;

469 H) and unlabelled (light; L) soil and leaf fractions were amplified with bacterial primers

470 341F and 785R [63]. Duplicate PCRs for each sample were pooled before purification

471 of PCR amplicons with a High Pure PCR product purification kit (Roche) according to

472 the manufacturer's instructions. It was not possible to obtain amplicons from ^{12}C-heavy

473 DNA (S 12C H) samples arising from SIP experiments with soils due to the low amount

474 of DNA present in these fractions (below the detection limit of 0.2 ng of the Qubit

475 dsDNA HS Assay kit). DNA libraries from purified 16S rRNA gene amplicons were

476 prepared and sequenced at MrDNA (Shallowater, TX, USA) with Illumina MiSeq

477 technology, obtaining an average of 100,757 16S rRNA genes per sample with an

478 average length of 300 bp. Sequence data were processed using MrDNA analysis

479 pipeline. Briefly, reads were first joined and depleted of barcodes. Then, short

480 sequences (<150 bp) and sequences with ambiguous base calls were removed.

481 Resultant sequences were denoised and operational taxonomic units (OTUs) were

482 defined with clustering at 97% similarity, followed by removal of singleton sequences

483 and chimeras. Taxonomy of OTUs was then assigned using BLASTn

484 (http://blast.ncbi.nlm.nih.gov) against a curated database derived from RPDII

485 (http://rdp/cme.msu.edu) and NCBI (www.ncbi.nlm.nih.gov).

486

487 <u>*isoA* gene amplicon sequencing</u>

488 *isoA* genes of DNA extracted from ^{13}C-heavy soil and leaf fractions were amplified with

489 primers isoA14F and isoA511R [35]. Duplicate PCRs for each replicate enriched with

490   ^{13}C-isoprene were pooled before purification of PCR amplicons with a High Pure PCR

491   product purification kit (Roche) as above. DNA libraries from purified PCR products

492   were prepared and sequenced at MrDNA (Shallowater, TX, USA) using Illumina MiSeq

493   technology, obtaining an average of 13,542 reads per sample with an average length

494   of 300 bp.

495   *isoA* amplicon sequencing data were analysed with the Bioconductor package DADA2

496   (version 1.6) [64] after demultiplexing the reads and removing primer sequences.

497   Reads were then trimmed to 275 nucleotides and quality-filtered if their expected error

498   was greater than two. Sequences were then denoised using the estimated error rates

499   and resultant reads were dereplicated. Subsequently, forward and reverse reads were

500   merged, chimeric sequences were discarded and the DADA2 algorithm was used to

501   infer individual amplicon sequence variants (ASVs). ASVs were then manually

502   checked by BLASTx [65]. Those ASVs with a top hit distinct from a ratified IsoA

503   sequence were discarded, obtaining a final set of 28 ASVs for soil and 6 ASVs for leaf

504   enrichments.

505

506   <u>Metagenomic analysis of oil palm soil and leaf DNA-SIP samples</u>

507   The 16S rRNA gene profile of each biological replicate of unenriched (T0), labelled

508   (heavy; H) and unlabelled (light; L) fractions from leaf and soil samples was analysed

509   by denaturing gradient gel electrophoresis (DGGE, see below). As they showed highly

510   similar profiles, biological replicates from both unenriched soils (S T0 R1, S T0 R2 and

511   S T0 R3) and phyllosphere (L T0 R1, L T0 R2 and L T0 R3) samples and ^{13}C-heavy

512   fractions from leaves (L 13C H R1, L 13C H R2 and L 13C H R3) were combined in

513   equal proportions for metagenomics sequencing (Fig S6), resulting in samples S T0,

514 L T0 and L 13C H R1-3, respectively. However, the community profile of ^{13}C-heavy

515 fraction from soils replicate 3 (S 13C H R3) showed some differences compared to

516 those from ^{13}C-heavy fractions from replicates 1 and 2 (S 13C H R1 and S 13C H R2;

517 Fig S6). Therefore, ^{13}C-heavy DNA soil fractions from replicates 1 and 2 were pooled

518 (S 13C H R1-2) for downstream analysis, whereas ^{13}C-heavy DNA from replicate 3 (S

519 13C H R3) was sequenced separately. Libraries were prepared by MrDNA

520 (Shallowater, TX, USA), resulting in an average library size of 700 bp for soils T0 (S

521 T0), 710 bp for leaves T0 (L T0), 643 bp for soils ^{13}C-heavy fractions from replicates

522 1 and 2 (S 13C H R1-2), 622 bp for soils ^{13}C-heavy fraction replicate 3 (S 13C H R3)

523 and 639 bp for ^{13}C-heavy leaves fractions (L 13C H R1-3). Libraries were then pooled

524 in equimolar ratios of 1 nM and sequenced as paired ends using the Illumina NovaSeq

525 6000 system.

526 Metagenomic reads were quality-filtered using the iu-filter-quality-minoche script [66]

527 included in Illumina-utils (version 1.4.4) [67], obtaining an average of 19,000,000

528 quality-filtered reads per sample with an average length of 143 bp. Taxonomy of

529 unassembled metagenomes was analysed using MetaPhlAn2 (version 2.0) [49].

530 The abundance of *isoA* genes in unassembled oil palm (this study), willow [32] and

531 poplar [38] unenriched metagenomes was determined by tBLASTn of IsoA sequences

532 from ratified isoprene-degrading bacteria against the raw reads (E ≤ 1e−4). Each

533 potential IsoA sequence retrieved from the analysis of metagenomes was manually

534 checked by BLASTx against a database of IsoA proteins from *bona fide* isoprene

535 degraders and discarded if they showed <50% amino acid identity. Only unique hits

536 were counted. Hit numbers were normalized against read number of the smallest

537 metagenome, to the smallest gene length and to hits of *recA*. Hits of *recA* in

538 unenriched metagenomes were determined by tBLASTn (E ≤ 1e-6) of RecA

539   sequences from a database obtained from RDP's FunGene [68]. Quality-filtered reads

540   from soil and leaf samples were assembled using metaSPAdes (version 3.13) [69]

541   with kmers 21, 33 and 55, and the quality of each assembly was analysed with

542   MetaQUAST (version 4.6.3) [70]. N50 values were ~1 kb for all metagenome

543   assemblies except for the unenriched soil (S T0) sample, which had an N50 value of

544   684. Complete statistics of metagenomes assemblies are shown in Table S2.

545   Assembled contigs were used to reconstruct metagenome assembled genomes

546   (MAGs) using MaxBin2 (version 2.2.2) [59]. MAGs completeness and contamination

547   was assessed and taxonomically assigned using CheckM (version 1.0.18) [60]. A total

548   of 20 MAGs from soil and 52 from leaf samples were obtained. Those MAGs with

549   >75% completeness and <5% contamination were then reassembled and

550   taxonomically verified using the "reassemble_bins" and "classify_bins" modules of the

551   metaWRAP pipeline (version 1.2.2) [71] to improve assembly and increase the

552   likelihood of obtaining full isoprene degradation gene clusters. Local BLAST

553   databases were constructed and screened for the presence of homologues to known

554   isoprene degradation proteins IsoABCDEFGHIJ and other polypeptides associated

555   with the pathway such as AldH1, GshB and GarB using a cut-off value of E<1e-10 in

556   permissive searches, and E<1e-40 in restrictive searches. A total of two MAGs from

557   soils and three from leaves were identified to be of interest based on a completion of

558   >75%, contamination <10% and the presence of IsoABCDEF homologues (Table S3).

559   Finally, MAGs were annotated using Prokka (version 1.13.3) [61].

560

561

562

Isolation and characterisation of isoprene-degrading bacteria

Once soils and leaf microcosms set up for DNA-SIP experiments had assimilated 50 µmol C per g of sample (10 days for leaves and 11 days for soils), they were diluted 1/10 into sealed 120 ml serum vials containing 10 ml Ewers minimal medium [62] to isolate isoprene-degrading strains from these environments. Vials were supplemented with 25 ppmv isoprene and incubated at 30 ˚C with shaking (150 rpm). Vials were subcultured three times at 2-week intervals before plating onto Ewers medium agar. These plates were incubated in air-tight jars containing isoprene vapour (1%, v/v) as sole carbon source. After four days, colonies with different morphologies were inoculated in 120 ml serum vials containing 10 ml Ewers medium and 25 ppmv isoprene to determine isoprene consumption of isolates in liquid medium. Consumption of isoprene was monitored using a gas chromatograph fitted with a flame ionization detector as previously described [33]. Those cultures that consistently consumed isoprene were checked for purity by plating onto rich medium R2A agar (Oxoid) and by phase contrast microscopy (Zeiss Axioscop).

Growth curves of selected isolates on isoprene were performed as follows: bacterial strains were first grown on Ewers medium with glucose (10 mM) for 48 h at 30 °C. Cultures were then pelleted and washed three times with Ewers medium with no carbon sources and adjusted to an OD_{600} of 0.8. Cell suspensions were then inoculated into 120 ml serum vials containing 20 ml of fresh Ewers medium amended with isoprene vapour (5%, v/v) as sole carbon source. Cultures were incubated at 30 °C and growth was estimated by measuring cell density at OD_{600} with a UV-1800 spectrophotometer (Shimadzu).

586 Genomic DNA of isoprene-degrading strains was extracted using a Wizard Genomic

587 DNA purification kit (Promega) according to the manufacturer's guidelines. The

588 phylogeny of isolates was identified by PCR amplification of their 16S rRNA genes

589 with primers 27F and 1492R [72] and subsequent Sanger sequencing (Eurofins

590 genomics). Genomic DNA was also examined for the presence of *isoA* genes using

591 PCR primers isoA14F and isoA511R, which are specific for the detection of *isoA*,

592 encoding the IsoMO α-subunit [35]. PCR conditions for amplification of *isoA* genes

593 consisted of an initial step of 94 °C for 2 min, followed by 31 cycles of 95 °C for 15 s,

594 54 °C for 30 s, 72 °C for 1 min and a final extension step of 72 °C for 7 min as described

595 in [35].

596

597 <u>Illumina and Nanopore sequencing of genomic DNA from *Variovorax* sp. OPL2.2</u>

598 Genomic DNA from *Variovorax* sp. OPL2.2 for Illumina and Nanopore sequencing was

599 extracted using a phenol: chloroform: isoamyl alcohol method described by Wilson

600 [73]. For Illumina sequencing, library preparation was done using the Nextera XT kit

601 (Illumina) according to the manufacturer's instructions. The library was checked for

602 quality and size distribution using the Agilent 2100 Bioanalyzer (Agilent) and the Qubit

603 dsDNA HS Assay Kit (Thermo Fisher Scientific). The library was then denatured and

604 sequenced with Illumina Miseq technology (San Diego, California USA). Paired-end

605 sequencing of 2 x 300 bp was performed using the MiSeq Reagent Kit v3 (San Diego,

606 California USA) according the manufacturer's protocol. Illumina sequencing of

607 *Variovorax* sp. OPL2.2 genomic DNA resulted in 1,274,620 reads.

608 Genomic DNA (0.9 µg) from *Variovorax* sp. OPL2.2 was used to prepare the library

609 for Nanopore sequencing. Library construction was performed using the Ligation

610 Sequencing Kit 1D (SQK-LSK109) in combination with the Native barcoding

611 Expansion Kit (EXP-NBD104) according the manufacturer's protocol (Oxford

612 Nanopore Technologies). DNA fragments were repaired and A-tailed using the

613 NEBNext® FFPE DNA Repair Mix and NEBNext® Ultra™ II End Repair/dA-Tailing

614 Module (New England Biolabs). After purification with AMPure XP beads (Beckman

615 Coulter Life Sciences), selected barcodes were ligated using the Blunt/TA Ligase

616 Master Mix (New England Biolabs). After ligation of the barcodes and bead clean-up,

617 the library was quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher

618 Scientific). Thereafter, adapters were ligated using the NEBNext® Quick Ligation

619 Module (New England Biolabs). The library was then purified and quantified with the

620 Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific) and sequenced using a Flow

621 Cell (R9.4.1) and MinION device (Oxford Nanopore Technologies) according to the

622 manufacturer's instructions.

623 Nanopore sequencing resulted in 44,530 reads (N50 = 10.5 kb). Base-calling and de-

624 multiplexing after sequencing was done using the Guppy Base-calling Software

625 (Oxford Nanopore Technologies, Limited Version 3.2.4+d9ed22f) resulting in 13,966

626 reads, selecting for a minimal sequence length of 3,000 bp.

627 Assembly of both Illumina and Nanopore reads to obtain a full-length genome was

628 done using Unicycler [74]. Assembly of *Variovorax* sp. OPL2.2 reads yielded 50

629 contigs with a total length of 8,516,444 bp, and the largest contig being 2,344,322 bp.

630 Completeness of the assembled genome was analysed by CheckM [60] resulting in

631 98.4% completeness and 1.1% contamination. Automatic annotation of *Variovorax* sp.

632 OPL2.2 genome was performed using Prokka [61].

633 Finally, average nucleotide identity (ANI) between *Variovorax* sp. OPL2.2 and

634 *Variovorax* sp. WS11 [32, 38] was calculated using the online ANI calculator available

635 on the Kostas Lab website (http://enve-omics.ce.gatech.edu/ani/index, accessed

636 10/12/2019).

637

638 <u>Denaturing gel gradient electrophoresis</u>

639 Bacterial 16S rRNA genes from DNA samples arising from DNA-SIP experiments were

640 amplified using primers 341F-GC [75] and 907R [76]. To visualise 16S rRNA gene

641 profiles of the bacterial communities from soil and leaf incubations, denaturing gel

642 electrophoresis (DGGE) was performed following the protocol described by El

643 Khawand *et al.* [36].

644

645 **Additional files**

646 **Additional file 1: Fig S1.** Bacterial diversity profile of the oil palm unenriched (T0)

647 community at the phyla level. **Fig S2.** Bacterial community profiles of oil palm soil

648 samples analysed by 16S rRNA gene amplicon sequencing. **Fig S3.** Bacterial

649 community composition of oil palm leaf samples analysed by 16S rRNA gene amplicon

650 sequencing. **Fig S4.** Growth curve of *Variovorax* sp. OPL2.2 on isoprene as sole

651 carbon and energy source. **Fig S5.** DNA retrieved as function of density of each

652 fraction recovered after isopycnic ultracentrifugation. **Fig S6.** 16S rRNA gene profiles

653 of oil palm soil and phyllosphere samples analysed by DGGE. **Table S1.** Relative

654 abundance of key isoprene-degrading bacterial genera in oil palm soil and leaf

655 samples. **Table S2.** Statistics for metagenome assemblies. **Table S3.** Metagenome

656 assembled genomes (MAGs) that contain genes encoding proteins homologous to

IsoABCDEF (E<1e-40). **Table S4.** MAGs genes encoding polypeptides homologous to proteins involved in isoprene metabolism from ratified isoprene-degrading strains. **Table S5.** ASVs retrieved from *isoA* amplicon sequencing analysis of ^{13}C-heavy DNA from soil and leaf incubations. **Table S6.** Location of oil palm trees used to set up soil and leaf DNA-SIP incubations (docx 2.0 Mb).

680   laboratory was funded by the UK Natural Environment Research Council (SUNRISE

681   project; NE/R000131/1).

688   **Authors' contributions**

689   OCF, LG, TJM and JCM planned the experiments. OCF and LG carried out the

690   experimental work and analysed results. LG designed and performed bioinformatics

691   analyses. DMOE, NPMcN and CVS collected oil palm soil and phyllosphere samples

692   in Malaysia. TAvA and HJMOdC performed the Illumina and Nanopore sequencing of

693   genomic DNA from *Variovorax* sp. OPL2.2. OCF, LG and JCM wrote the manuscript

694   with contributions from all authors. All authors read and approved the manuscript

695   before submission.

697   **Ethics approval and consent to participate**

698   Not applicable

700   **Consent for publication**

701   Not applicable

Competing interests

The authors declare that they have no competing interests.

References

1. Atkinson R, Arey J. Gas-phase tropospheric chemistry of biogenic volatile organic compounds: A review. Atmos Environ. 2003;37:197-219.

2. Guenther AB, Jiang X, Heald CL, Sakulyanontvittaya T, Duhl T, Emmons LK, *et al*. The model of emissions of gases and aerosols from nature version 2.1 (MEGAN2.1): An extended and updated framework for modeling biogenic emissions. Geosci Model Dev. 2012;5:1471–92.

3. United Nations Statistics Division: The sustainable development goals report 2019. https://unstats.un.org/sdgs/report/2019 (2019). Accessed 6 Mar 2020.

4. Griggs D, Stafford-Smith M, Gaffney O, Rockström J, Öhman MC, Shyamsundar P, *et al*. Sustainable development goals for people and planet. Nature. 2013;495:305-7.

5. Monson RK, Trahan N, Rosenstiel TN, Veres P, Moore D, Wilkinson M, *et al*. Isoprene emission from terrestrial ecosystems in response to global change: minding the gap between models and observations. Philos Trans A Math Phys Eng Sci. 2007;365:1677-95.

6. Sanderson MG, Jones CD, Collins WJ, Johnson CE, Derwent RG. Effect on climate change on isoprene emissions and surface ozone levels. Geophys Res Lett. 2003;30:1936.

7. Engelhart GJ, Moore RH, Nenes A, Pandis SN. Cloud condensation nuclei activity of isoprene secondary organic aerosol. J Geophys Res. 2011;116:D02207.

8. Morais ARC, Dworakowska S, Reis A, Gouveia L, Matos CT, Bogdal D, *et al*. Chemical and biological-based isoprene production: Green metrics. Catal Today. 2015;239:38-43.

9. Dani SKG, Benavides AMS, Michelozzi M, Peluso G, Torzillo G, Loreto F. Relationship between isoprene emission and photosynthesis in diatoms, and its implications for global marine isoprene estimates. Mar Chem. 2017;189:17–24.

10. Geron C, Owen S, Guenther A, Greenberg A, Rasmussen R, Hui Ban J, *et al*. Volatile organic compounds from vegetation in southern Yunnan Province, China: Emission rates and some potential regional implications. Atmos Environ. 2006;40:1759-73.

730   11. Bäck J, Aaltonen H, Hellén H, Kajos MK, Patotoski J, Taipale R, *et al.* Variable emissions of
731   microbial volatile organic compounds (MVOCs) from root-associated fungi isolated from Scots
732   pine. Atmos Environ. 2010;44:3651-9.

733   12. Exton DA, Suggett DJ, McGenity TJ, Steinke, M. Chlorophyll-normalized isoprene production
734   in laboratory cultures of marine microalgae and implications for global models. Limnol
735   Oceanogr. 2013;58:1301–11.

736   13. Fall R, Copley SD. Bacterial sources and sinks of isoprene, a reactive atmospheric
737   hydrocarbon. Environ Microbiol. 2000;2:123–130.

738   14. Kuzma J, Nemecek-Marshall M, Pollock WH, Fall R. Bacteria produce the volatile hydrocarbon
739   isoprene. Curr Microbiol. 1995;30:97–103.

740   15. Shaw SL, Gantt B, Meskhidze N. Production and emissions of marine isoprene and
741   monoterpenes: a review. Adv Meterol. 2010:408696.

742   16. Steinke M, Hodapp B, Subhan R, Bell TG, Martin-Creuzburg D. Flux of the biogenic volatiles
743   isoprene and dimethyl sulphide from an oligotrophic lake. Sci Rep. 2018;8:630.

744   17. Logan BA, Monson RK, Potosnak MJ. Biochemistry and physiology of foliar isoprene
745   production. Trends Plant Sci. 2000;5:477–81.

746   18. Loreto F, Ciccioli P, Brancaleoni E, Valentini R, De Lillis M, Csiky O *et al.* A hypothesis on the
747   evolution of isoprenoid emission by oaks based on the correlation between emission type and
748   *Quercus* taxonomoy. Oeacologia. 1998;115:17430-35.

749   19. Monson RK, Jones RT, Rosenstiel TN, Schnitzler JP. Why only some plants emit isoprene.
750   Plant Cell Environ. 2013;36:503-16.

751   20. Sharkey TD. Is it useful to ask why plants emit isoprene? Plant Cell Environ. 2013;36:517-20.

752   21. Rohmer, M. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis
753   in bacteria, algae and higher plants. Nat Prod Rep. 1999;16:565-74.

754   22. Monson RK, Fall R. Isoprene emission from aspen leaves: influence of environment and
755   relation to photosynthesis and photorespiration. Plant Physiol. 1989;90:267-74.

756   23. Sharkey TD, Yeh S. Isoprene emission from plants. Annu Rev Plant Physiol Plant Mol Biol.
757   2001;52:407-36.

758   24. Behnke K, Ehlting B, Teuber M, Bauerfeind M, Louis S, Hänsch R *et al.* Transgenic, non-
759   isoprene emitting plants don't like it hot. Plant J. 2007;51:485-99.

760   25. Laothawornkitkul J, Paul ND, Vickers CE, Possell M, Taylor JE, Mullineaux PM *et al*. Isoprene

761   emissions influence herbivore feeding decisions. Plant Cell Environ. 2008;31:1410-5.

762   26. Vickers CE, Possell M, Cojocariu CI, Velikova VB, Laothawornkitkul J, Ryan A *et al*. Isoprene

763   synthesis protects transgenic tobacco plants from oxidative stress. Plant Cell Environ. 2009;32-

764   520-31.

765   27. Lantz AT, Allman J, Weraduwage SM, Sharkey TD. Isoprene: New insights into the control of

766   emission and mediation of stress tolerance by gene expression. Plant Cell Environ.

767   2019;42:2808-26.

768   28. Cleveland CC, Yavitt JB. Consumption of atmospheric isoprene in soil. Geophys Res Lett.

769   1997;24-2379-82.

770   29. Cleveland CC, Yavitt JB. Microbial consumption of atmospheric isoprene in a temperate forest

771   soil. Appl Env Microbiol. 1998;64:172–77.

772   30. Gray CM, Helmig D, Fierer N. Bacteria and fungi associated with isoprene consumption in soil.

773   Elem Sci Anth. 2015;3:000053.

774   31. McGenity TJ, Crombie AT, Murrell JC. Microbial cycling of isoprene, the most abundantly

775   produced biological volatile organic compound on Earth. ISME J. 2018;12:931-41.

776   32. Larke-Mejía NL, Crombie AT, Pratscher J, McGenity TJ, Murrell JC. Novel isoprene-degrading

777   proteobacteria from soil and leaves identified by cultivation and metagenomics analysis of

778   stable isotope probing experiments. Front Microbiol. 2019;10:2700.

779   33. Crombie AT, Khawand ME, Rhodius VA, Fengler KA, Miller MC, Whited GM *et al.* Regulation

780   of plasmid-encoded isoprene metabolism in *Rhodococcus*, a representative of an important link

781   in the global isoprene cycle. Environ Microbiol. 2015;17:3314-29.

782   34. Leahy JG, Batchelor PJ, Morcomb SM. Evolution of the soluble diiron monooxygenases. FEMS

783   Microbiol Rev. 2003;27:449-479.

784   35. Carrión O, Larke-Mejía NL, Gibson L, Farhan Ul Haque M, Ramiro-García J, McGenity TJ, *et*

785   *al*. Gene probing reveals the widespread distribution, diversity and abundance of isoprene-

786   degrading bacteria in the environment. Microbiome. 2018;6:219.

787   36. El Khawand M, Crombie AT, Johnston A, Vavlline DV, McAuliffe JC, Latone JA *et al.* Isolation

788   of isoprene degrading bacteria from soils, development of *iso*A gene probes and identification

789   of the active isoprene-degrading soil community using DNA-stable isotope probing. Environ

790   Microbiol; 2016;18:2743–53.

791 37. Neufeld JD, Vohra J, Dumont MG, Lueders T, Manefield M, Friedrich MW, *et al.* DNA stable-

792   isotope probing. Nat Protoc. 2007;2:860–6.

793 38. Crombie AT, Larke-Mejía NL, Emery H, Dawson R, Pratscher J, Murphy GP, *et al.* Poplar

794   phyllosphere harbors disparate isoprene-degrading bacteria. Proc Natl Acad Sci USA.

795   2018;115:13081-6.

796 39. Kesselmeier J, Staudt M. Biogenic volatile organic compounds (VOC): An overview on

797   emission, physiology and ecology. J Atmos Chem. 1999;33:23-88.

798 40. Cheng Y, Yu L, Xu Y, Lu H, Cracknell AP, Kanniah K *et al.* Mapping oil palm plantation

799   expansion in Malaysia over the past decade (2007-2016) using ALOS-1/2 PALSAR-1/2 data.

800   Int J Remote Sens. 2019;40:7389-408.

801 41. Ibragimov A, Sidique SF, Tey YS. Productivity for sustainable growth in Malaysian oil palm

802   production: a system dynamics modeling approach. J Clean Prod. 2019;213:1051–62.

803 42. Hewitt CN, MacKenzie AR, Di Carlo P, Di Marco CF, Dorsey JR, Evans M *et al.* Nitrogen

804   management is essential to prevent tropical oil palm plantations from causing ground-level

805   ozone pollution. Proc Natl Acad Sci USA. 2009;106:18447-51.

806 43. Fulthorpe RR, Roesch LF, Riva A, Triplett EW. Distantly sampled soils carry few species in

807   common. ISME J. 2008;2:901-10.

808 44. Janssen PH. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA

809   genes. Appl Environ Microbiol. 2006;72:1719-28.

810 45. Karimi B, Terrat S, Dequiedt S, Saby NPA, Horrigue W, Lelièvre M, *et al.* Biogeography of soil

811   bacteria and archaea across France. Sci Adv. 2018;4:eaat1808.

812 46. Roesch LF, Fulthorpe RR, Riva A, Casella G, Hadwin AK, Kent AD, *et al.* Pyrosequencing

813   enumerates and contrasts soil microbial diversity. ISME J. 2007;1:283-90.

814 47. Acuña Alvarez LA, Exton DA, Timmis KN, Suggett DJ, McGenity TJ. Characterization of marine

815   isoprene-degrading communities. Environ Microbiol. 2009;11:3280–91.

816 48. Johnston A, Crombie AT, Khawand ME, Sims L, Whited GM, McGenity TJ *et al.* Identification

817   and characterisation of isoprene-degrading bacteria in an estuarine environment. Environ

818   Microbiol. 2017;19:3526–37.

819   49. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, *et al*. MetaPhlAn2 for

820   enhanced metagenomic taxonomic profiling. Nat Methods. 2015;12:902-3.

821   50. Kembel SW, O'Connor TK, Arnold HK, Hubbell SP, Wright SJ, *et al*. Relationships between

822   phyllosphere bacterial communities and plant functional traits in a neotropical forest. Proc Natl

823   Acad Sci USA. 2014;111:13715-20.

824   51. Lambais MR, Crowley DE, Cury JC, Büll RC, Rodrigues RR. Bacterial diversity in tree canopies

825   of the Atlantic forest. Science. 2006;312:1917.

826   52. Redford AJ, Bowers RM, Knight R, Linhart Y, Fierer N. The ecology of the phyllosphere:

827   geographic and phylogenetic variability in the distribution of bacteria on tree leaves. Environ

828   Microbiol. 2010;12:2885-93.

829   53. Vorholt JA. Microbial life in the phyllosphere. Nat Rev Microbiol. 2012;10:828-40.

830   54. Knief C, Delmotte N, Chaffron S, Stark M, Innerebner G, Wassmann R, von Mering C, Vorholt

831   JA. Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere

832   of rice. ISME J. 2012;6:1378-90.

833   55. Rastogi G, Sbodio A, Tech JJ, Suslow TV, Coaker GL, Leveau JHJ. Leaf microbiota in an

834   agroecosystem: spatiotemporal variation in bacterial community composition on field-grown

835   lettuce. ISME J. 2012;6:1812-22.

836   56. Bodenhausen N, Horton MW, Bergelson J. Bacterial communities associated with the leaves

837   and the roots of *Arabidopsis thaliana*. PLoS One. 2013;8:e56329.

838   57. Wei YJ, Wu Y, Yan YZ, Zou W, Xue J, Ma WR, *et al*. High-throughput sequencing of microbial

839   community diversity in soil, grapes, leaves, grape juice and wine of grapevine from China. PLoS

840   One. 2018;13:e0193097.

841   58. Rasmussen RA, Khalil MAK. Isoprene over the Amazon basin. J Geophys Res-Atmos.

842   1988;93:1417-21.

843   59. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover

844   genomes from multiple metagenomic datasets. Bioinformatics. 2016;32:605-7.

845   60. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the

846   quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome

847   Res. 2015;25:1043-55.

848   61. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068-9.

849   62. Dorn E, Hellwig M, Reineke W, Knackmuss HJ. Isolation and characterization of a 3-
850   chlorobenzoate degrading pseudomonad. Arch Microbiol. 1974;99:61–70.

851   63. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. Evaluation of
852   general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-
853   based diversity studies. Nucleic Acids Res. 2013;41:e1.

854   64. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: high-
855   resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581-3.

856   65. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, *et al.* BLAST+:
857   architecture and applications. BMC Bioinformatics. 2009;10:421.

858   66. Minoche AE, Dohm JC, Himmelbauer H. Evaluation of genomic high-throughput sequencing
859   data generated on Illumina HiSeq and genome analyzer systems. Genome Biol. 2011;12:R112.

860   67. Eren AM, Vineis JH, Morrison HG, Sogin ML. A filtering method to generate high quality short
861   reads using illumina paired-end technology. PLoS One. 2013;8:e66643.

862   68. Fish JA, Chai B, Wang Q, Sun Y, Brown CT, Tiedje JM, *et al.* FunGene: the functional gene
863   pipeline and repository. Front Microbiol. 2013;4:291.

864   69. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile
865   metagenomics assembler. Genome Res. 2017;27:824-34.

866   70. Mikheenko A, Saveliev V, Gurevich A. MetaQUAST: evaluation of metagenome assemblies.
867   Bioinformatics. 2016;32:1088-90.

868   71. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP-a flexible pipeline for genome-resolved
869   metagenomic data analysis. Microbiome. 2018;6:158.

870   72. Lane DJ. 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M, editors. Nucleic acid
871   techniques in bacterial systematics. Chichester: John Wiley & Sons; 1991. p. 115-75.

872   73. Wilson K. Preparation of genomic DNA from bacteria. Curr Protocol Mol Biol. 2001;56:2.4.1-
873   2.45.

874   74. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from
875   short and long sequencing reads. PLoS Comput Biol. 2017;13:e1005595.

876   75. Muyzer G, de Waal EC, and Uitterlinden, AG. Profiling of complex microbial populations by
877   denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes
878   coding for 16S rRNA. Appl Environ Microbiol. 1993;59: 695–700.

879   76. Muyzer G, Brinkhoff T, Nuebel U, Santegoeds C, Schäefer H, Waver C. Denaturing Gradient

880   Gel Electrophoresis (DGGE) in microbial ecology. In: Akkermans ADL, van Elsas JD, de Bruijim

881   FJ, editors. Molecular microbial ecology manual. Dordrecht: Kluwer Academic Publishers;

882   1998. p. 1-27.

883

884

885   **Figure legends**

886

887   **Fig 1. Bacterial community profile of oil palm soil samples.** Bacterial diversity of the unenriched

888   (T0) and labelled (heavy; H) fractions of ^{13}C-isoprene soil incubations was analysed by 16S rRNA gene

889   amplicon (16S) and metagenomics (MG) sequencing. The unenriched bacterial community (S T0)

890   represents the average of three biological replicates. ^{13}C-heavy DNA from replicates 1 and 2 were

891   combined before MG sequencing due to their highly similar 16S rRNA gene community profile shown

892   by DGGE, resulting in sample S 13C H R1-2 (see Methods). Only genera with >5 RA% in at least one

893   of the conditions are represented. Genera present at >10% in any sample are shown in bold. Genera

894   with <5% RA are grouped in "Others". For complete 16S rRNA gene amplicon sequencing data,

895   including individual replicates and ^{12}C-isoprene controls, see Fig S2.

896

897   **Fig 2. Bacterial community composition of oil palm phyllosphere samples.** Bacterial diversity

898   profile of unenriched (T0) and labelled (heavy; H) fractions of ^{13}C-isoprene incubations of oil palm leaf

899   samples was analysed by 16S rRNA gene amplicon (16S) and metagenomics (MG) sequencing. The

900   unenriched bacterial community (L T0) represents the average of three biological replicates. ^{13}C-heavy

901   DNA from replicates 1, 2 and 3 of leaf incubations were combined before MG sequencing due to their

902   highly similar 16S rRNA gene community profile shown by DGGE, resulting in sample L 13C H R1-3

903   (see Methods). Only genera with >5 RA% in at least one of the conditions are represented. Genera

904   present at >10% in any sample are shown in bold. Genera with <5% RA are grouped in "Others". For

905 complete 16S rRNA gene amplicon sequencing data, including individual replicates and ^{12}C-isoprene

906 controls, see Fig S3.

907

908 **Fig 3. Isoprene metabolic gene clusters from representative isoprene-degrading strains (in bold)**

909 **and metagenome assembled genomes (MAGs).** Genes encoding IsoMO (*isoABCDEF*) are coloured

910 in red. Adjacent genes not involved in isoprene degradation are coloured in white. Regulatory genes

911 are shown in black. "\\" represents a discontinuity between two DNA contigs. *Variovorax* sp. OPL2.2

912 was isolated in this study from oil palm leaf enrichments.

913

914 **Fig 4. Relative abundance and diversity of *isoA* genes in ^{13}C-heavy fractions from oil palm**

915 **incubations.** *isoA* sequences in ^{13}C-heavy fractions from oil palm soil (S 13C H) and leaf (L 13C H)

916 samples were analysed by *isoA* amplicon sequencing (see Methods). Only AVSs with >5% RA in at

917 least one replicate are represented. ASVs present at >10% in any sample are shown in bold. ASVs

918 closely related to IsoA from *Rhizobiales* MAG are represented in purple; ASVs with highest homology

919 to IsoA from *Novosphingobium* MAG are shown in green; ASVs closely related to IsoA from

920 *Rhodococcus* are coloured in blue; ASVs with highest homology to IsoA from *Gordonia* MAG are

921 represented in orange. ASVs with RA <5% are grouped as "Others". For complete *isoA* gene amplicon

922 sequencing data, see Table S5.

923

Additional File 1

Diversity of isoprene-degrading bacteria in phyllosphere and soil communities from a high isoprene-emitting environment: a Malaysian oil palm plantation

Ornella Carrión[1‡*], Lisa Gibson[1‡], Dafydd M.O. Elias[2], Niall P. McNamara[2], Theo A. van Alen[3], Huub J.M. Op den Camp[3], Christina Vimala Supramaniam[4], Terry J. McGenity[5], J. Colin Murrell[1*]

[1]School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK

[2]Centre of Ecology and Hydrology, Lancaster University, Bailrigg, Lancaster, LA1 4AP, UK

[3]Department of Microbiology, Faculty of Science, IWWR, Radboud University Nijmegen, Heyendaalseweg 135, NL-6525 AJ Nijmegen, The Netherlands

[4]School of Biosciences, Nottingham Centre of Sustainable Palm Oil, University of Nottingham-Malaysia, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia

[5]School of Life Sciences, University of Essex, Colchester, UK

[‡]These authors contributed equally to this work

*Corresponding authors:

J Colin Murrell, [1]School of Environmental Sciences, University of East Anglia

Norwich Research Park, NR4 7TJ, UK

E-mail: j.c.murrell@uea.ac.uk

Tel: (+44) 01603 592959

Ornella Carrión, [1]School of Environmental Sciences, University of East Anglia

Norwich Research Park, NR4 7TJ, UK

E-mail: o.carrion-fonseca@uea.ac.uk

Tel: (+44) 01603 592239
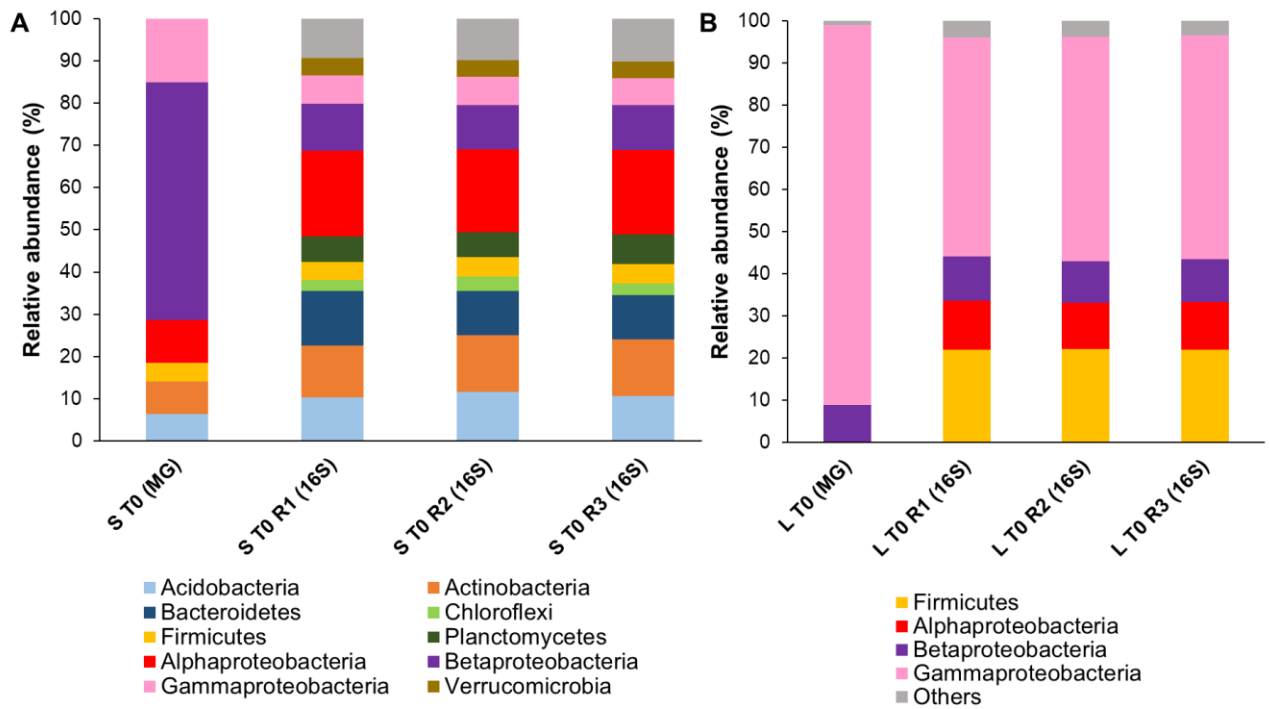
Fig S1. Bacterial diversity profile of the oil palm unenriched (T0) community at the phyla level.

Bacterial community composition was analysed by 16S rRNA gene amplicon (16S) and metagenomics (MG) sequencing. A: soil T0 samples; B: phyllosphere T0 samples. Only phyla with >2% relative abundance in at least one of the samples are represented. The Proteobacteria phylum is subdivided into Alpha-, Beta- and Gammaproteobacteria classes. Bacterial phyla with <2% relative abundance are recorded as "Others".

Fig S2. Bacterial community profiles of oil palm soil samples analysed by 16S rRNA gene amplicon sequencing. T0: unenriched samples; 13C UF: unfractionated DNA from ^{13}C-isoprene enrichments; 13C H: heavy DNA from ^{13}C-isoprene incubations; 13C L: light DNA from ^{13}C-isoprene samples; 12C L: light DNA from ^{12}C-isoprene control samples. Only OTUs with >5% relative abundance in at least one of the replicates are represented. Genera present at >10% in any sample are shown in bold. OTUs with <5% relative abundance are shown as "Others".
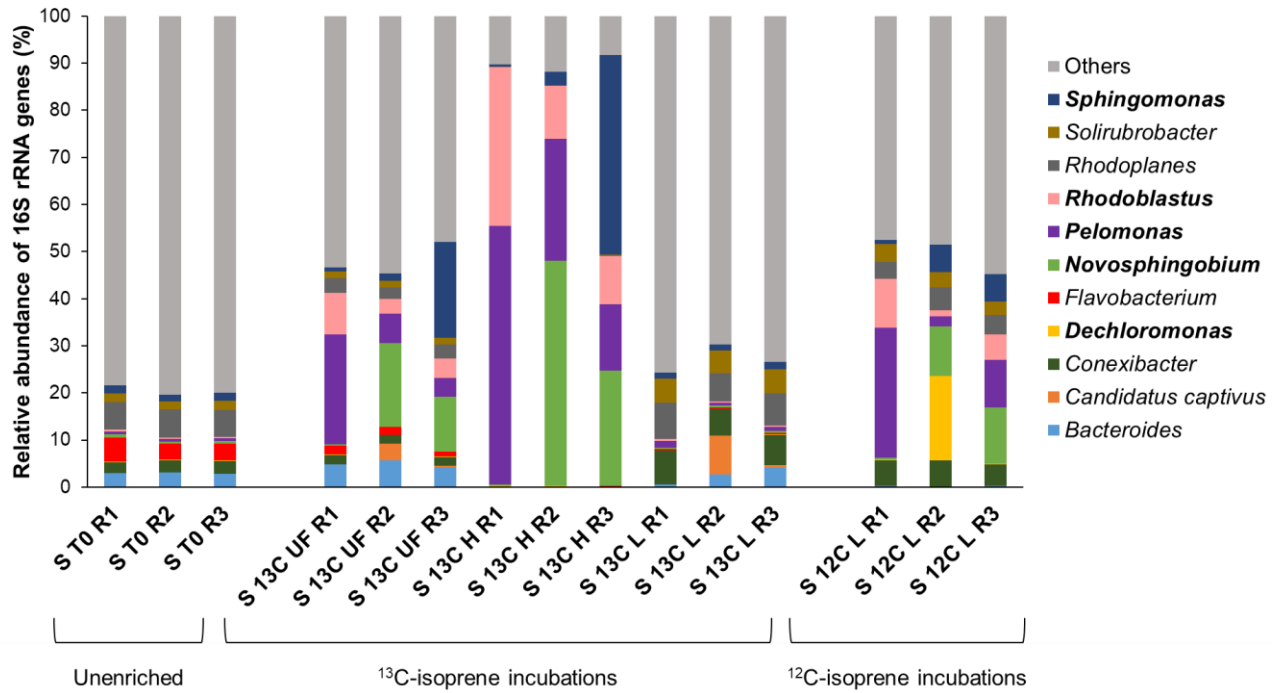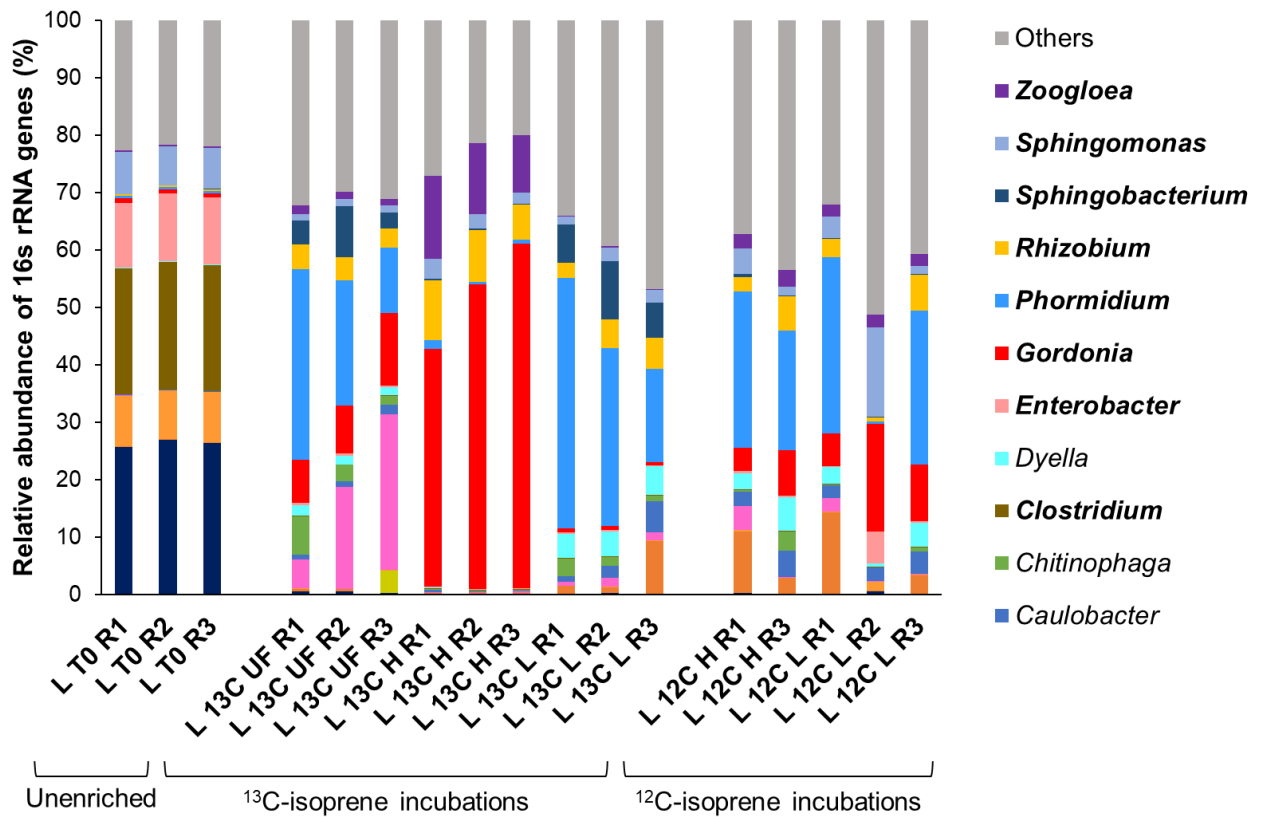
Fig S3. Bacterial community composition of oil palm leaf samples analysed by 16S rRNA gene amplicon sequencing. T0: unenriched samples; 13C UF: unfractionated DNA from ^{13}C-isoprene enrichments; 13C H: heavy DNA from ^{13}C-isoprene incubations; 13C L: light DNA from ^{13}C-isoprene samples; 12C H: heavy DNA from ^{12}C-isoprene control incubations; 12C L: light DNA from ^{12}C-isoprene control samples. Only OTUs with >5% relative abundance in at least one of the replicates are represented. Genera present at >10% in any sample are shown in bold. OTUs with <5% relative abundance are recorded as "Others".
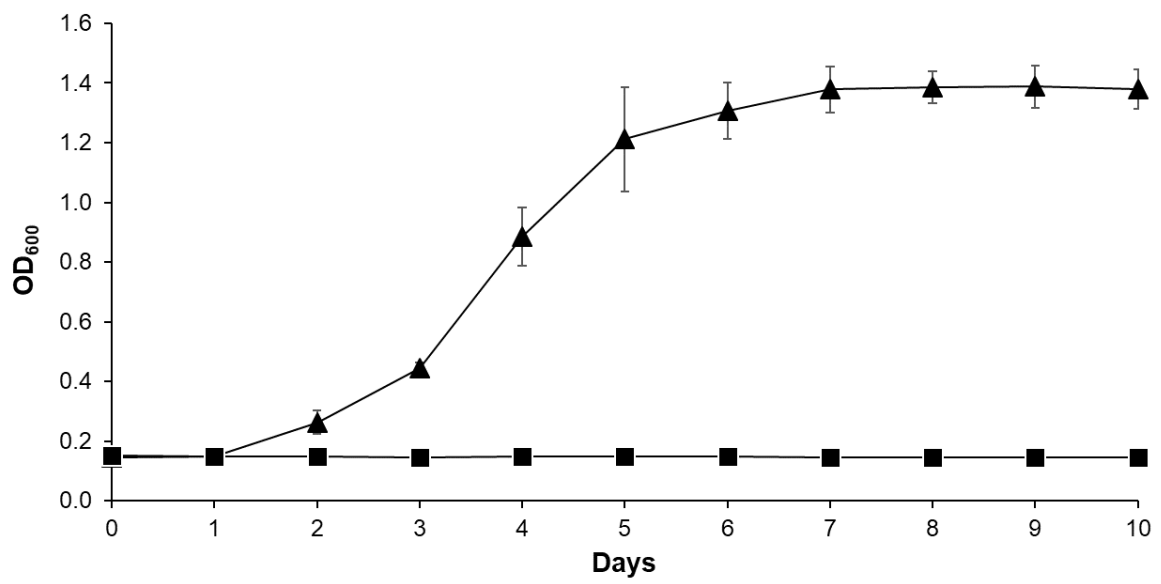
Fig S4. Growth curve of *Variovorax* sp. OPL2.2 on isoprene as sole carbon and energy source.
Triangles: *Variovorax* sp. OPL2.2 cultures amended with isoprene (5%; v/v). Data points represent the average of three biological replicates with their respective standard deviations. Squares: *Variovorax* sp. OPL2.2 culture with no carbon source added (control; one replicate).
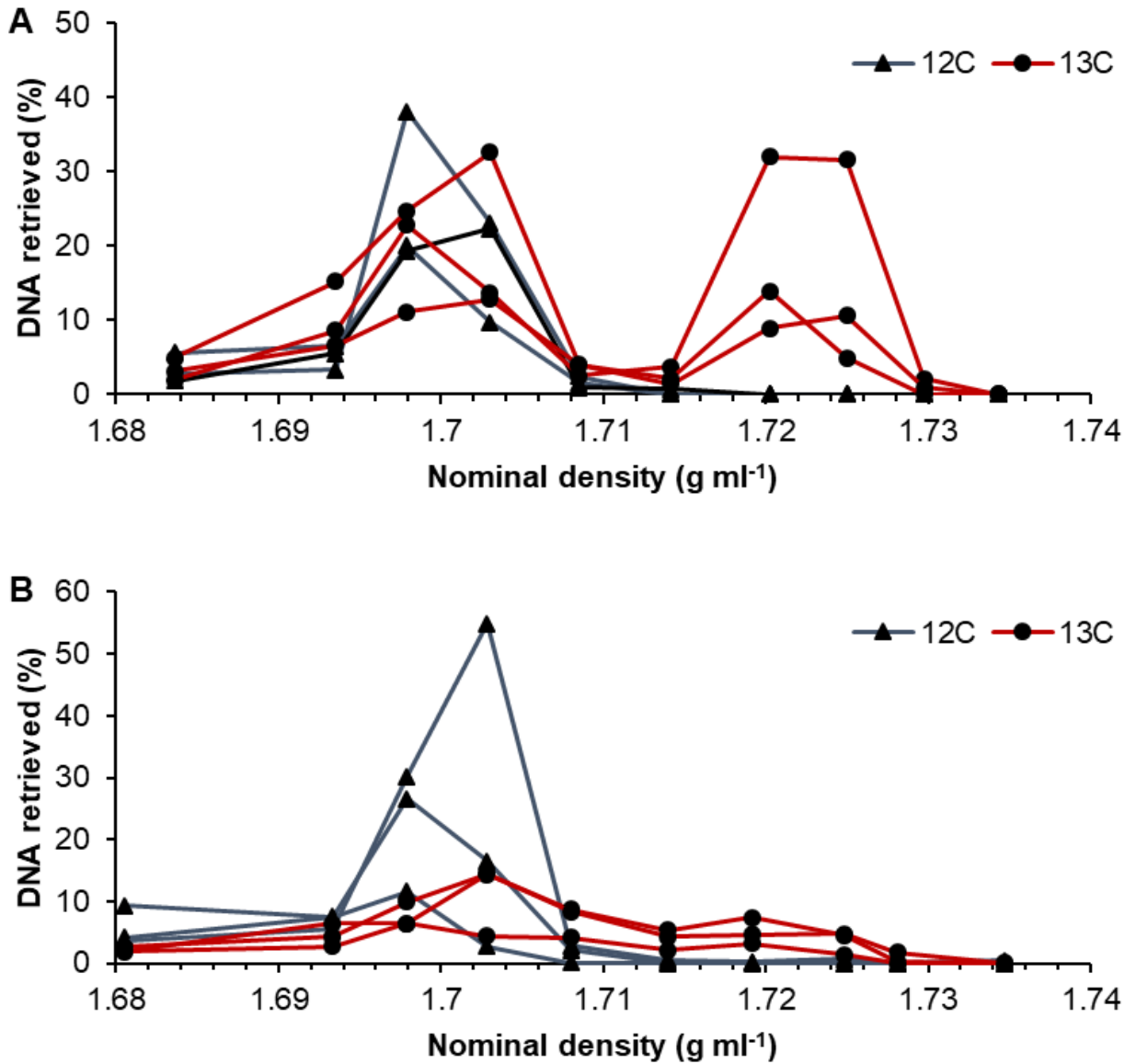
Fig S5. DNA retrieved as function of density of each fraction recovered after isopycnic ultracentrifugation. A: soil samples incubated with ^{12}C (blue line) or ^{13}C-isoprene (red line); B: leaf incubations with ^{12}C (blue line) or ^{13}C-isoprene (red line).
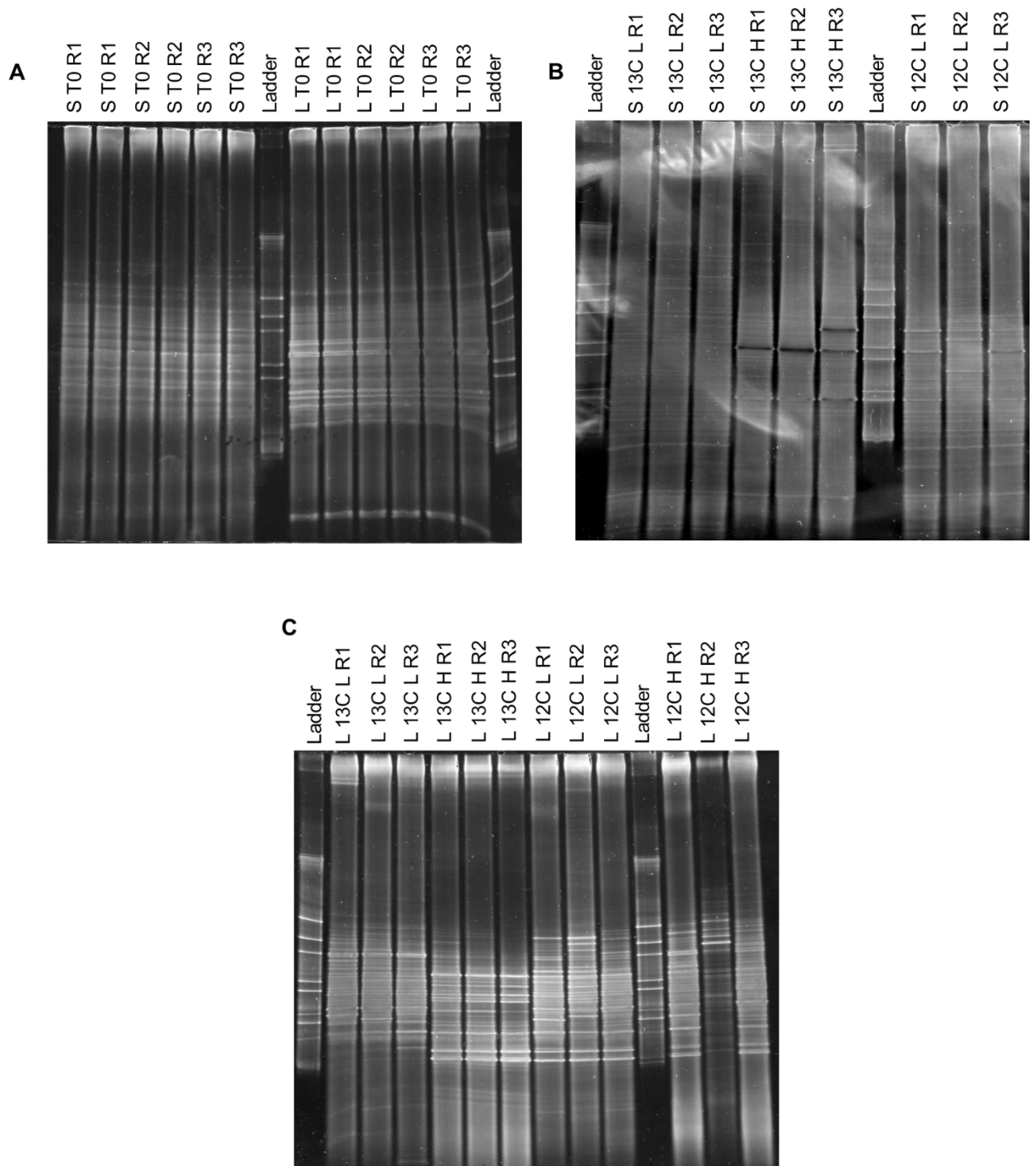
Fig S6. 16S rRNA gene profiles of oil palm soil and phyllosphere samples analysed by DGGE. A: biological replicates from unenriched soil (S T0) and phyllosphere (L T0) samples, each of which was ran in duplicate. B: biological replicates from heavy (H) and light (L) soil incubations with ^{13}C or ^{12}C-isoprene (control). C: biological replicates from heavy (H) and light (L) leaf samples enriched ^{13}C or ^{12}C-isoprene.

Table S1. Relative abundance of key isoprene-degrading bacterial genera in oil palm soil and leaf samples.

Genus	S T0	L T0	S 13C H	L 13C H
Gordonia	0.7 ± 0.2%	0.8 ± 0.1%	0.9 ± 0.1%	51.4 ± 9.3%
Novosphingobium	0.6 ± 0.1%	2.8 ± 0.1%	24.2 ± 23.8%	0.3 ± 0.1%
Pelomonas	0.5 ± <0.1%	0.4 ± 0.1%	31.7 ± 21.0%	0.4 ± 0.1%
Rhodoblastus	0.4 ± 0.1%	0.2 ± <0.1%	18.4 ± 13.3%	0.3 ± 0.1%
Sphingomonas	1.5 ± 0.1%	7.1 ± 0.3%	15.2 ± 23.6%	2.6 ± 0.7%
Zoogloea	0.2 ± <0.1%	0.3 ± 0.1%	0.3 ± 0.1%	12.3 ± 2.2%

Relative abundance of bacterial genera in oil palm incubations was analysed by 16S rRNA gene amplicon sequencing. Values shown represent the average of three biological replicates with their respective standard deviations. S T0: unenriched soil samples; L T0: unenriched leaf samples; S 13C H: heavy DNA from soil samples enriched with ^{13}C-isoprene; L 13C H: heavy DNA from leaf samples enriched with ^{13}C-isoprene.

Table S2. Statistics for metagenome assemblies.

	S T0	S 13C H R1-2	S 13C H R3	L T0	L 13C H R1-3
Contigs	138,886	36,166	22,852	76,131	114,878
Largest contig (bp)	75,971	703,868	1,215,493	732,750	621,426
Total length (bp)	101,440,201	64,831,442	49,671,873	152,876,094	150,393,192
GC (%)	63.9	64.9	64.6	56.5	66
N50	684	2,852	5,677	5,018	1,565
N75	568	1,078	1,322	1,171	798
L50	50,076	3,502	959	3,953	18,953
L75	91,069	13,317	6,174	21,847	54,079
Bins	4	9	7	28	24

Metagenomes from unenriched soil (S T0) and leaf (L T0) samples and ^{13}C-heavy fractions from soil (S 13C H R1-2 and S 13C H R3) and leaf (L 13C H R1-3) incubations were assembled with metaSPAdes using kmers 21, 33 and 55. Quality of the assemblies was analysed by MetaQUAST and bins obtained using MaxBin2.

Table S3. Metagenome assembled genomes (MAGs) that contain genes encoding proteins homologous to IsoABCDEF (E<1e-40).

MAG	Metagenome of origin	Size (Mbp)	N50	Completeness (%)	Contamination (%)	Strain heterogeneity (%)
Novosphingobium	S 13C H R3	3.7	441,005	99.5	<0.1	0
Rhizobiales	S 13C H R3	4.1	59,990	97.6	2.5	79
Gordonia polyisoprenovorans	L 13C H R1-3	6.1	194,257	99.8	0.9	40
Zoogloeaceae	L 13C H R1-3	5.2	49,423	98.7	2.1	10
Ralstonia	L 13C H R1-3	4.5	5,039	79	6.4	0

MAGs were reconstructed from metagenomic sequencing of ^{13}C-heavy DNA from soil (S 13C H) and leaf (L 13C H) incubations with ^{13}C-isoprene. MAGs completeness and contamination was assessed and taxonomically assigned using CheckM (see Methods). N50 is calculated for contigs.

Table S4. MAGs genes encoding polypeptides homologous to proteins involved in isoprene metabolism from ratified isoprene-degrading strains.

Novosphingobium MAG

Gene	Description	Closest protein from ratified isoprene degrader	Amino acid identity (%)	Coverage (%)
isoA	Hydroxylase α-subunit	IsoA from *Sphingopyxis* sp. OPL5	95.2	100
isoB	Hydroxylase γ-subunit	IsoB from *Sphingopyxis* sp. OPL5	100	85.1
isoC	Reiske-type ferrodoxin	IsoC from *Sphingopyxis* sp. OPL5	100	78.6
isoD	Coupling protein	IsoD from *Sphingopyxis* sp. OPL5	100	90.6
isoE	Hydroxylase β-subunit	IsoE from *Sphingopyxis* sp. OPL5	83.7	100
isoF	Flavoprotein NADH reductase	IsoF from *Sphingopyxis* sp. OPL5	76.2	100
isoG	Racemase	IsoG from *Sphingopyxis* sp. OPL5	91.1	100
isoH	Dehydrogenase	IsoH from *Sphingopyxis* sp. OPL5	89.4	100
isoI	Glutathione-S-transferase	IsoI from *Sphingopyxis* sp. OPL5	89.8	100
isoJ	Glutathione-S-transferase	IsoJ from *Sphingopyxis* sp. OPL5	90.6	98
aldH1	Aldehyde dehydrogenase	AldH1 from *Sphingopyxis* sp. OPL5	85.3	100

Rhizobiales MAG

Gene	Description	Closest protein from ratified isoprene degrader	Amino acid identity (%)	Coverage (%)
isoA	Hydroxylase α-subunit	IsoA from *Sphingopyxis* sp. OPL5	84.9	100
isoB	Hydroxylase γ-subunit	IsoB from *Sphingopyxis* sp. OPL5	67.4	97
isoC	Reiske-type ferrodoxin	IsoC from *Sphingopyxis* sp. OPL5	61.2	92
isoD	Coupling protein	IsoD from *Ramlibacter* sp. WS9	60	92
isoE	Hydroxylase β-subunit	IsoE from *Sphingopyxis* sp. OPL5	54.8	98
isoF	Flavoprotein NADH reductase	IsoF from *Sphingopyxis* sp. OPL5	54.9	99
isoG	Racemase	IsoG from *Sphingopyxis* sp. OPL5	72.9	96
isoH	Dehydrogenase	IsoH from *Sphingopyxis* sp. OPL5	74.3	100
isoI	Glutathione-S-transferase	IsoI from *Sphingopyxis* sp. OPL5	72.9	95
isoJ	Glutathione-S-transferase	IsoJ from *Sphingopyxis* sp. OPL5	69.9	100
aldH1	Aldehyde dehydrogenase	AldH1 from *Sphingopyxis* sp. OPL5	64.7	99

Gordonia polyisoprenovorans MAG

Gene	Description	Closest protein from ratified isoprene degrader	Amino acid identity (%)	Coverage (%)
isoA	Hydroxylase α-subunit	IsoA from *Gordonia polyisoprenovorans* i37	98.8	100
isoB	Hydroxylase γ-subunit	IsoB from *Gordonia polyisoprenovorans* i37	82.1	100
isoC	Reiske-type ferrodoxin	IsoC from *Gordonia polyisoprenovorans* i37	99.1	100
isoD	Coupling protein	IsoD from *Ramlibacter* sp. WS9	99	95
isoE	Hydroxylase β-subunit	IsoE from *Gordonia polyisoprenovorans* i37	98.8	100
isoF	Flavoprotein NADH reductase	IsoF from *Gordonia polyisoprenovorans* i37	99.4	100
isoG	Racemase	IsoG from *Gordonia polyisoprenovorans* i37	98.3	100
isoH	Dehydrogenase	IsoH from *Gordonia polyisoprenovorans* i37	99.1	100
isoI	Glutathione-S-transferase	IsoI from *Gordonia polyisoprenovorans* i37	100	100
isoJ	Glutathione-S-transferase	IsoJ from *Gordonia polyisoprenovorans* i37	94.7	100
aldH1	Aldehyde dehydrogenase	AldH1 from *Gordonia polyisoprenovorans* i37	83	100
aldH2	Aldehyde dehydrogenase	AldH2 from *Gordonia polyisoprenovorans* i37	94.8	100
gshB	Glutathione synthetase	GshB from *Gordonia polyisoprenovorans* i37	99.1	98
coADR	CoA disulfide reductase	CoADR from *Gordonia polyisoprenovorans* i37	94	100

Zoogloeaceae MAG

Gene	Description	Closest protein from ratified isoprene degrader	Amino acid identity (%)	Coverage (%)
isoA	Hydroxylase α-subunit	IsoA from *Variovorax* sp. WS11	51.1	100
isoB	Hydroxylase γ-subunit	IsoB from *Ramlibacter* sp. WS9	51.7	97
isoC	Reiske-type ferrodoxin	IsoC from *Ramlibacter* sp. WS9	51	93
isoD	Coupling protein	IsoD from *Gordonia* sp. OPL2	44.9	85
isoE	Hydroxylase β-subunit	IsoE from *Ramlibacter* sp. WS9	43	85
isoF	Flavoprotein NADH reductase	IsoF from *Ramlibacter* sp. WS9	39	97

Ralstonia MAG

Gene	Description	Closest protein from ratified isoprene degrader	Amino acid identity (%)	Coverage (%)
isoA	Hydroxylase α-subunit	IsoA from *Variovorax* sp. WS11	48.3	100
isoB	Hydroxylase γ-subunit	IsoB from *Ramlibacter* sp. WS9	52.5	83
isoC	Reiske-type ferrodoxin	IsoC from *Ramlibacter* sp. WS9	47.6	92
isoD	Coupling protein	IsoD from *Rhodococcus opacus* PD630	44.2	81
isoE	Hydroxylase β-subunit	IsoE from *Nocardioides* sp. WS12	45.1	94
isoF	Flavoprotein NADH reductase	IsoF from *Rhodococcus opacus* PD630	39.1	100

Homology of the polypeptides encoded by genes recovered from MAGs to proteins from ratified isoprene degraders was analysed by BLASTx (see Methods) and is expressed as amino acid identity.

Table S5. ASVs retrieved from *isoA* amplicon sequencing analysis of ^{13}C-heavy DNA from soil and leaf incubations.

	Closest IsoA sequence	Source	Amino acid Identity(%)	Cover (%)	Soil Samples RA(%)	Leaf Samples RA(%)
ASV_1	*Gordonia i37*	Isolate	100	99	5.3±4.6	91.9±7.3
ASV_2	*Rhizobiales*	Metagenome	100	99	35.5±6.1	
ASV_4	*Rhizobiales*	Metagenome	92.5	99	6.1±5.3	
ASV_5	*Rhizobiales*	Metagenome	100	99	8.9±8.7	
ASV_7	*Rhizobiales*	Metagenome	100	99	4.2±3.7	
ASV_8	*Rhizobiales*	Metagenome	71	99	2.2±3.8	
ASV_9	*Rhizobiales*	Metagenome	98.7	99	1.8±3.1	
ASV_11	*Novosphingobium*	Metagenome	100	99	3.7±3.4	
ASV_12	*Rhizobiales*	Metagenome	98.9	99	1.4±1.3	
ASV_13	*Rhodococcus* AD45	Isolate	100	99	3.9±3.4	0.9±1.3
ASV_14	*Rhizobiales*	Metagenome	98.1	99	1.8±0.4	
ASV_15	*Rhizobiales*	Metagenome	96.2	99	2.5±2.4	
ASV_16	*Rhodococcus* AD45	Isolate	84.91	99	2.9±2.1	
ASV_17	*Rhizobiales*	Metagenome	92.45	99	0.6±1	
ASV_18	*Rhizobiales*	Metagenome	98.9	99	2.6±4.5	
ASV_19	*Rhizobiales*	Metagenome	100	99	1.8±3.1	
ASV_21	*Rhizobiales*	Metagenome	98.7	99	0.3±0.6	
ASV_23	*Rhodococcus* AD45	Isolate	99.37	99	1.2±1	
ASV_24	*Rhizobiales*	Metagenome	100	99	0.2±0.3	
ASV_26	*Rhizobiales*	Metagenome	98.1	99	0.2±0.4	
ASV_29	*Rhizobiales*	Metagenome	100	99	1.8±3.2	
ASV_30	*Rhizobiales*	Metagenome	98.7	99	0.1±0.3	
ASV_34	*Gordonia i37*	Isolate	98.1	99		3.4±5.6
ASV_37	*Rhizobiales*	Metagenome	96.2	99	0.1±0.2	
ASV_44	*Novosphingobium*	Metagenome	99.4	99	2.6±4.6	
ASV_53	*Rhizobiales*	Isolate	99.4	95	0.5±0.8	
ASV_59	*Gordonia i37*	Isolate	99.4	96		1.8±3
ASV_85	*Gordonia* i37	Isolate	99.4	96		2.7±3.8
ASV_110	*Rhizobiales*	Metagenome	99.4	100	0.1±0.2	
ASV_114	*Gordonia i37*	Isolate	100	99		0.9±1.3
ASV_122	*Rhizobiales*	Metagenome	95.6	100	0.1±0.2	
ASV_160	*Novosphingobium*	Metagenome	100	99	0.2±0.4	
ASV_168	*Gordonia i37*	Isolate	93.1	100		1.7±3.0

Amino acid identity of ASVs retrieved from soil (S 13C H) and phyllosphere (L 13C H) samples to IsoA from ratified isoprene-degrading strains or MAGs reconstructed from metagenomes analysed in this study was determined by BLASTx (see Methods). Relative abundance (RA) of each ASV in heavy fractions from ^{13}C-isoprene soil and leaf incubations represent the average of three biological replicates with their respective standard deviations. ND: not detected.

Table S6. Location of oil palm trees used to set up soil and leaf DNA-SIP incubations.

Sample	Latitude (N)	Longitude (E)	Elevation (m)
Palong A	2° 54' 03"	102° 39' 54"	55
Palong B	2° 54' 06"	102° 39' 33"	63
Palong C	2° 54' 03"	102° 38' 56"	33

Soil and leaf samples were collected from the same trees to allow comparison of the diversity of isoprene degraders from both environments.
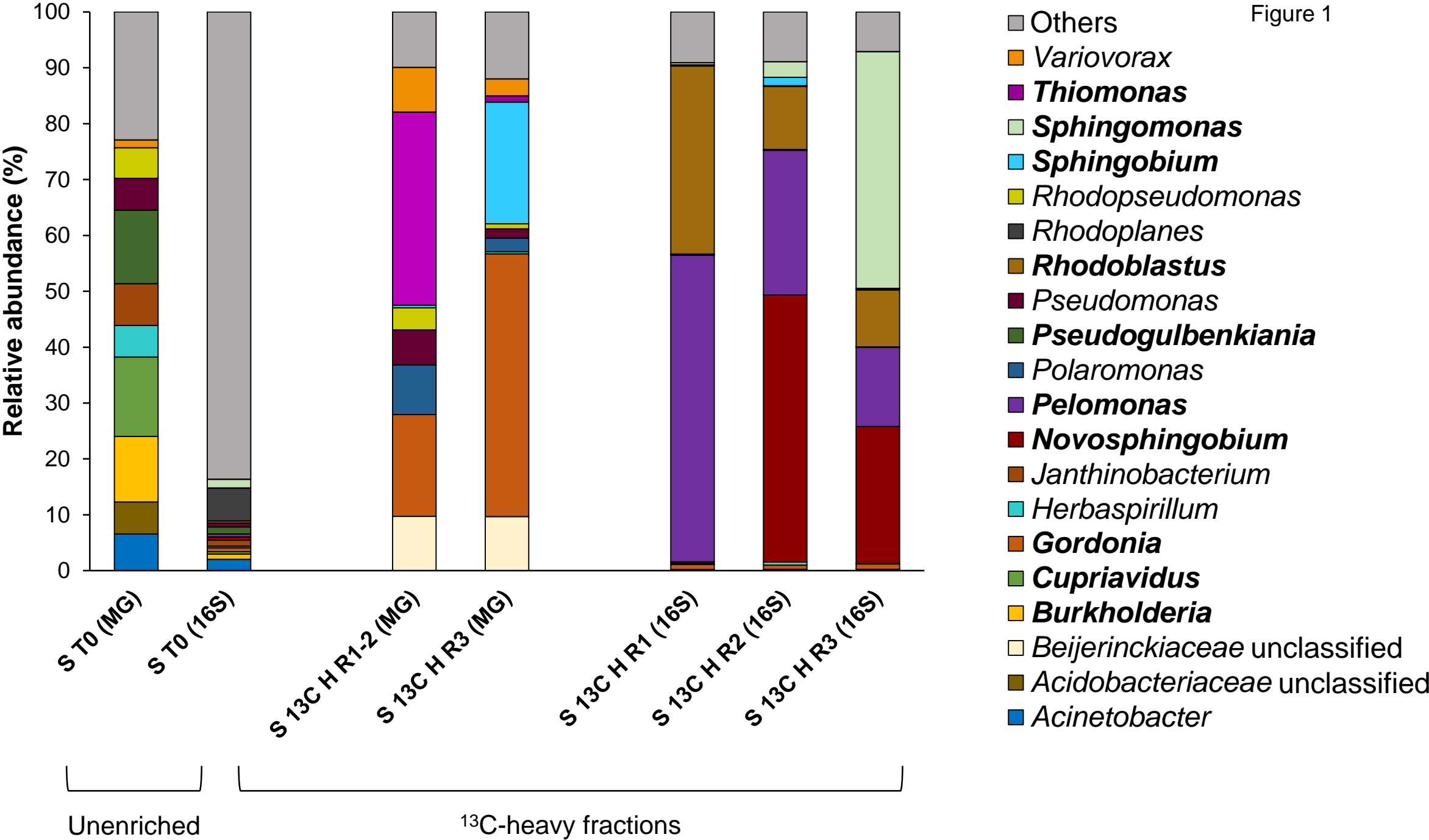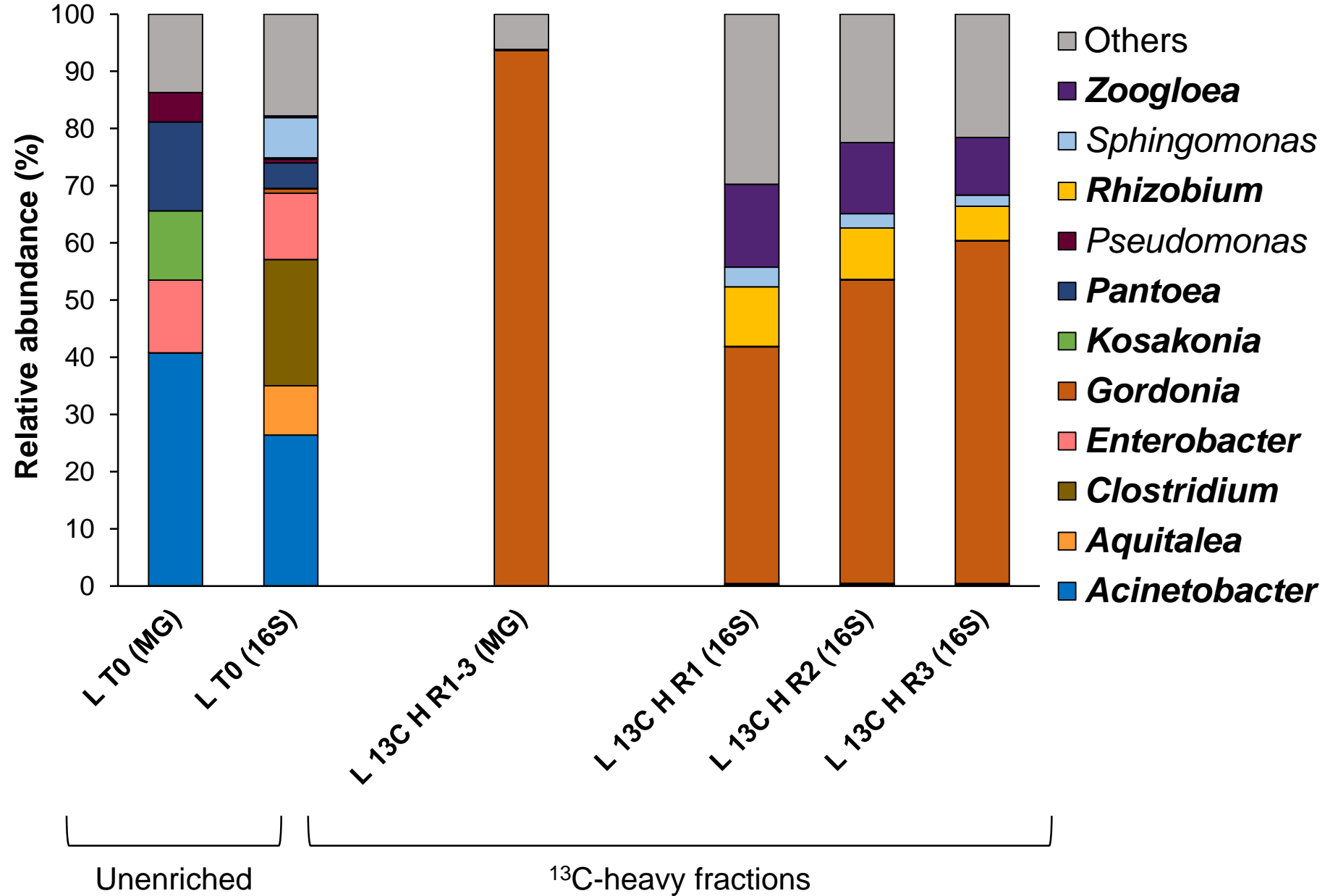
Figure 1

Legend (top to bottom):
- Others
- *Variovorax*
- ***Thiomonas***
- ***Sphingomonas***
- ***Sphingobium***
- *Rhodopseudomonas*
- *Rhodoplanes*
- ***Rhodoblastus***
- *Pseudomonas*
- ***Pseudogulbenkiania***
- *Polaromonas*
- ***Pelomonas***
- ***Novosphingobium***
- *Janthinobacterium*
- *Herbaspirillum*
- ***Gordonia***
- ***Cupriavidus***
- ***Burkholderia***
- *Beijerinckiaceae* unclassified
- *Acidobacteriaceae* unclassified
- *Acinetobacter*

X-axis labels: S T0 (MG), S T0 (16S), S 13C H R1-2 (MG), S 13C H R3 (MG), S 13C H R1 (16S), S 13C H R2 (16S), S 13C H R3 (16S)

Y-axis: Relative abundance (%)

Brackets: Unenriched; ^{13}C-heavy fractions

Figure 2

Gordonia sp. i37
isoG2 *isoH2* *marR* *gshA* *isoG* *isoH* *isoI* *isoJ* *aldh1* *isoA* *isoBCD* *isoE* *isoF* *aldH2* *gshB* *CoA-DSR*

Gordonia (MAG)

Novosphingobium (MAG)

Rhizobiales (MAG)

Ralstonia (MAG)

Zoogloeaceae (MAG)

Variovorax sp. OPL2.2
dmlR *garB* *dmlR*

Variovorax sp. WS11
dmlR *dmlR*

Figure 3

Figure 4