

Single-cell genomics based on Raman sorting reveals novel carotenoid-containing bacteria in the Red Sea

Yizhi Song,¹ Anne-Kristin Kaster,² John Vollmers,² Yanqing Song,³ Paul A. Davison,⁴ Martinique Frentrup,² Gail M. Preston,⁵ Ian P. Thompson,¹ J. Colin Murrell,⁶ Huabing Yin,³ C. Neil Hunter⁴ and Wei E. Huang^{1,*}

¹Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK.

²Leibniz Institute DSMZ, Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Inhoffenstrasse 7 B, 38124, Braunschweig, Germany.

³Division of Biomedical Engineering, School of Engineering, University of Glasgow, Glasgow G12 8QQ, UK.

⁴Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN, UK.

⁵Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK.

⁶School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

Summary

Cell sorting coupled with single-cell genomics is a powerful tool to circumvent cultivation of microorganisms and reveal microbial ‘dark matter’. Single-cell Raman spectra (SCRSs) are label-free biochemical ‘fingerprints’ of individual cells, which can link the sorted cells to their phenotypic information and ecological functions. We employed a novel Raman-activated cell ejection (RACE) approach to sort single bacterial cells from a water sample in the Red Sea based on SCRS. Carotenoids are highly diverse pigments and play an important role in phototrophic bacteria, giving strong and distinctive Raman spectra. Here, we showed that individual carotenoid-containing cells from a Red Sea sample were isolated based on the characteristic SCRS. RACE-based single-cell

genomics revealed putative novel functional genes related to carotenoid and isoprenoid biosynthesis, as well as previously unknown phototrophic microorganisms including an unculturable *Cyanobacteria* spp. The potential of Raman sorting coupled to single-cell genomics has been demonstrated.

Introduction

Microbes are the most diverse and abundant organisms on earth and play critical roles in biogeochemical carbon and nitrogen cycling (Paterson *et al.*, 1997; Whitman *et al.*, 1998; Schleifer, 2004; Huang *et al.*, 2009a,b). However, the vast majority of microbes in natural environments have not yet been grown in the laboratory using traditional cultivation methods (Amann *et al.*, 1995; Venter, 2003; Venter *et al.*, 2004; Daniel, 2005; Swan *et al.*, 2011; Rinke *et al.*, 2013; Hedlund *et al.*, 2014). Furthermore, to understand the function of microbes in a community, it is desirable to place even cultivable microorganisms in their ecological and environmental context instead of just studying pure cultures (Huang *et al.*, 2015).

Metagenomics is a powerful cultivation-independent approach for studying microbes. It can provide an overview of the diversity and metabolic blueprints of potential functions of microbes in environmental samples. Unfortunately, assembling individual discrete genomes from metagenomics data is quite difficult and rather costly, especially for samples with a high diversity of microorganisms. Although binning algorithms, which group contigs and assign DNA sequences to operational taxonomic units, have massively improved over the years, metagenomes assembled in this way may still be mosaics of DNA from different strains. Single-cell genomics can complement metagenomics approaches and be used to establish genetic linkages between DNA sequences within individual cells (Swan *et al.*, 2011; Lasken, 2012; Rinke *et al.*, 2013). Although fluorescent-activated cell sorting has been successfully used to sort single cells and subsequently perform single-cell genomics (Rinke *et al.*, 2013), Raman-activated cell sorting (RACS) can provide a valuable alternative to sort cells based on single-cell Raman spectra (SCRSs) which can reflect phenotypic and intrinsic biochemical fingerprints of cells (Huang *et al.*, 2004, 2015). An SCRS is able to provide a label-free biochemical profile of a single cell, which could contain information on nucleic acids, proteins, carbohydrates, lipids within the cell and specific

Received 14 July, 2016; revised 19 July, 2016; accepted 14 September, 2016. *For correspondence. E-mail wei.huang@eng.ox.ac.uk; Tel. +44 1865 283786; Fax +44 (0)1865 374992.

Microbial Biotechnology (2016) 0(0), 000–000
doi:10.1111/1751-7915.12420

Funding Information

WEH acknowledges support from EPSRC (EP/M002403/1 and EP/M02833X/1) and NERC (NE/M002934/1) and CNH thanks BBSRC sLoLa (BB/M000265/1) in the UK for the financial support. JCM acknowledges funding from the Gordon and Betty Moore Foundation Marine Microbiology Initiative Grant #3303.

© 2016 The Authors. *Microbial Biotechnology* published by John Wiley & Sons Ltd and Society for Applied Microbiology.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Raman-active compounds (e.g. carotenoids, isoprenoid, cytochrome *c*, poly- β -hydroxybutyrate and glycogen) (Li *et al.*, 2012a,b,c; Zhang *et al.*, 2015a,b). RACS-mediated single-cell genomics should, therefore, help to reveal phenotypic information and couple certain phenotypes to genotypes of cells.

Raman-activated cell sorting can be achieved by combining Raman single-cell detection with optical tweezers (Huang *et al.*, 2009a,b), microfluidic devices (Zhang *et al.*, 2015a,b), or the so-called Raman-activated cell ejection (RACE) (Wang *et al.*, 2013). Optical tweezers and microfluidic-based RACS methods are, however, quite labour intensive and difficult to apply to complex samples. Previously, RACE has been performed by using two separate instruments: Raman and laser microdissection microscopes (Wang *et al.*, 2013). In this study, we integrated Raman single-cell detection and cell ejection into one system. This new system is now able to characterize single cells based on SCRS and accurately isolate cells of interest on a slide surface by employing laser-induced forward

transfer (LIFT) (Hopp *et al.*, 2005). LIFT applies a laser to a thin-layer-coated surface which causes immediate and *local* evaporation of the coated layer, therefore pushing or ablating the selected cell on the layer into a collection device (Fig. 1).

Carotenoids are one of the most diverse chemicals in bacteria (Takaichi, 2008) and give strong and distinctive Raman spectra (Kochendoerfer *et al.*, 1999; Krebs *et al.*, 2003; Robert, 2009; Li *et al.*, 2012a,b,c). Due to the resonance Raman effect, the Raman spectra of carotenoids are sharp and strong, providing typical and unambiguous carotenoid bands (Kochendoerfer *et al.*, 1999; Krebs *et al.*, 2003; Robert, 2009) making them ideal targets for the RACE system. Carotenoids are biologically important molecules, present in nearly all photosynthetic cells (Li *et al.*, 2012a,b,c). For example, carotenoids are associated with both chlorophyll-based photochemical reaction centres and rhodopsin-based light-activated proton pumps (Bryant and Frigaard, 2006), which are two important light-harvesting (LH) systems of

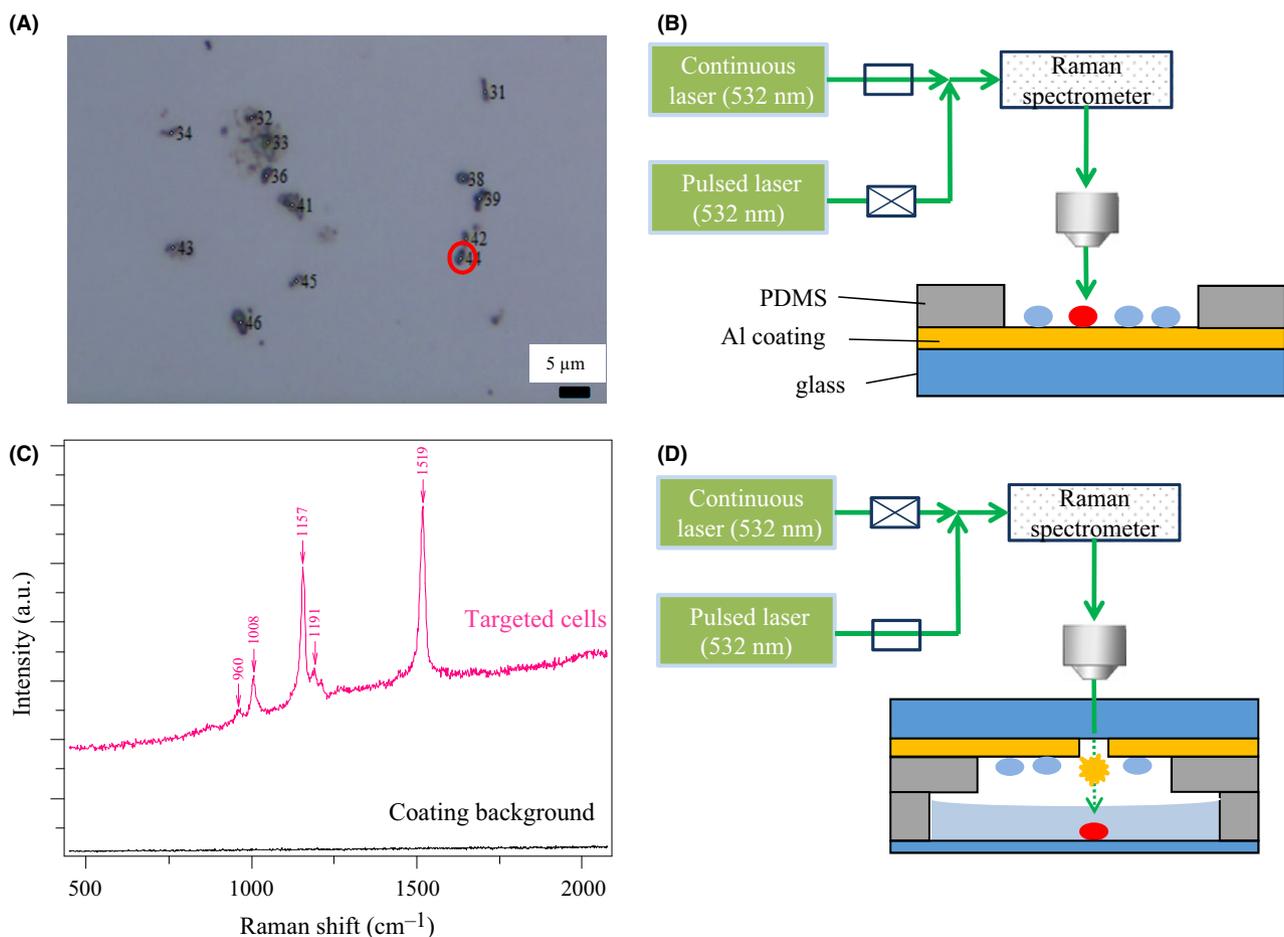


Fig. 1. Illustration of resonant Raman-activated single-cell ejection (RACE). (A) Microscopic image of cells on a RACE chip. (B) The continuous laser is used for acquiring single-cell Raman spectra. PDMS: Polydimethylsiloxane (C) Spectra of one single cell and the slide coating background. (D) The RACE chip is turned over and the target cell is ejected into the collector by a pulsed laser.

photosynthesis. In chlorophyll-based LH systems, carotenoids absorb solar energy, quench free radicals and are structurally important parts of chlorophyll-based LH complexes (Garcia-Asua *et al.*, 1998). In rhodopsin-based LH systems, a carotenoid, namely β -carotene, is a precursor of retinal, which is an essential component of bacteriorhodopsin and proteorhodopsin (Gonzalez *et al.*, 2011). Hence, carotenoids are important indicative molecules related to light harvesting. Sorting carotenoid-containing cells would, therefore, help to probe novel uncultured phototrophic bacteria. In this study, RACE was used to isolate carotenoid-containing bacteria from the Red Sea. Both new photosynthetic bacteria and putative novel functional genes were discovered.

Results

Establishing the RACE system by integrating a cell ejection system with a Raman micro-spectroscope

A 532 nm pulsed laser for single-cell ejection has been integrated into a Raman micro-spectroscope, which is designated as RACE (Fig. 1). Cells were added onto a RACE sampling chip (Fig. 1A and Fig. S1) and examined by the Raman micro-spectroscope using a continuous 532 nm laser (Fig. 1B). SCRSs were used as sorting criteria to distinguish the phenotypic 'profile' of cells (in this case, the presence of carotenoids), and the positions of targeted cells were recorded (Fig. 1C). The RACE sampling chip was turned over, and single cells of interest were then ejected and harvested in a RACE collection chip (Fig. 1D and Fig. S2).

Multiple displacement amplification from single cells isolated from RACE

Genomes of single or multiple (3–8) cells sorted with the RACE system were amplified by multiple displacement amplification (MDA; Blanco *et al.*, 1989). As an initial test, *Escherichia coli* JM109 cells containing the plasmid p18GFP were used to examine the performance of single-cell ejection. The green fluorescent protein gene *gfp* on the plasmid p18GFP and the universal stress protein A gene *uspA* (Chen and Griffiths, 1998) on the chromosome were used to assess gene recovery. The *gfp* on this plasmid has about 20–40 copies per cell, whereas *uspA* is a single copy gene on the *E. coli* chromosome. Figure S3A shows that the negative control did not contain amplified DNA, and three of seven samples contained MDA products from single *E. coli* cells. The other four samples did not yield visible MDA products (the products were too little to be seen in the gel), which may be due to the inefficient MDA of these cells. The PCR product of the *gfp* gene was recovered from all seven samples, while only one of the seven samples contained

the *uspA* PCR product (Fig. S3B and S3C). This is probably due to amplification bias in the MDA, since multiple-copy *gfp* genes have a better chance of being amplified than a single-copy gene such as *uspA* in the one cell. Although sometimes the MDA yield was not high enough to be clearly viewed in agarose gel, the amount of DNA obtained was sufficient to perform PCR amplification (Fig. S3B and S3C). Nonetheless, this shows that it is possible that a single-copy functional gene can be recovered by MDA from a single-cell isolated by RACE.

Sorting carotenoid-containing bacteria from a Red Sea water sample based on their resonant Raman spectra

Raman-activated cell ejection has been applied to sort bacterial cells sampled from surface seawater at Eilat on the Red Sea coast. SCRS of 5321 single cells have been obtained and analysed, and 33% of the analysed cells exhibited strong fluorescence and 67% had distinguishable SCRS for sorting. Figure 2 shows the principal component analysis (PCA) of those spectra, and the spots within the red line area are the cells containing carotenoids according to Raman spectra. The most significant loadings of PCA at axis 1 are 997–1007, 1145–1161 and 1503–1526 cm^{-1} (Fig. S4), corresponding to characteristic ν_1 , ν_2 and ν_3 Raman bands of carotenoids (Robert, 2009; Li *et al.*, 2012a,b,c). This is in a good agreement with the fact that carotenoid-containing cells are distributed along axis 1 within the red line area (Fig. 2).

Figure 3A shows a few examples of SCRS, including a typical bacterial cell, individual cells showing

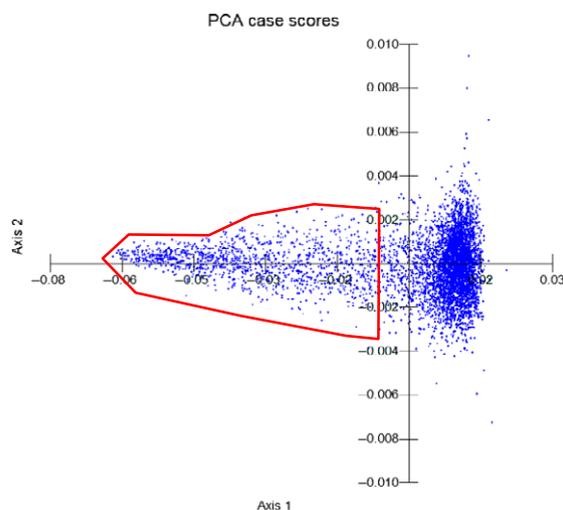


Fig. 2. The PCA analysis of 5321 SCRS from the Red Sea sample. The group within red line area is cells containing various carotenoids, and the rest of cells have no carotenoids.

fluorescence, cells containing poly- β -hydroxybutyrate (PHB) and unidentified compounds (cell images are shown in Fig. S5A). Since SCRS can reflect the biochemical phenotypes of cells (Huang *et al.*, 2004, 2007a,b, 2009a,b; Li *et al.*, 2012a,b,c; Berry *et al.*, 2015), cells can be sorted based on the Raman biomarker bands of SCRS. For example, cells containing PHB have distinguishable Raman biomarkers (Majed and Gu, 2010) at 839, 1058, 1403 and 1123 cm^{-1} (Fig. 3A), which could be used to sort PHB containing cells. Interestingly, some SCRS also indicate that cells contain unidentified novel compounds, e.g. Fig. 3A shows a SCRS of a cell with a hydroxyisoquinoline-like compound (Fig. S5A panel v). Since RACS is a new technology, a database with well-characterized biomarkers is needed for the future analysis of cells.

In this study, the characteristic ν_1 , ν_2 and ν_3 Raman bands of carotenoids were used as sorting criteria (Fig. 3). The ν_1 band is the methyl rocking mode, and the ν_2 and ν_3 bands vary due to the different lengths of conjugated C=C bonds and stretching modes of C-C bonds (Robert, 2009) in different carotenoids. Since the Raman spectral resolution is about 1 cm^{-1} , variations in the positions of ν_1 , ν_2 and ν_3 indicate the different structures of carotenoid presented in the cells. According to SCRS, 1223 (23%) cells contained carotenoids. Based on their carotenoid SCRS, 30 attempts of single-cell ejection and whole genome amplification (WGA) were performed. Seven of the 30 attempts were successful in which 16S rRNA was recovered from the genome amplification. The SCRS of these seven successful cases are shown in Fig. 3B and 3C, and their information are listed in Table 1. Collectively these seven sorted samples could be resolved into five different types of carotenoids. Optical images of the sorted cells show that the cells had size of 0.3–1 μm (Fig. S5B).

Genome sequencing of isolated cells

The estimated degree of genome completeness and putative contamination was based on universal marker genes analyses. The overall contamination based on CheckM analyses (Parks *et al.*, 2015) was estimated to be relatively low in all cases (Table S3). However, sequence analyses also indicated relatively low genome coverage – less than 20% in all cases. The genome sequence data revealed functional genes related to carotenoid biosynthesis, CO_2 fixation, haem/chlorophyll biosynthesis and fatty acid, alcohol and aldehyde metabolism. These data are summarized in Table 1. A detailed list of functional genes is given in Table S4. Despite the low estimated genome coverage, genes associated with carotenoid biosynthesis were identified in six of the seven sorted samples (P728-5, B728-3, H808-5,

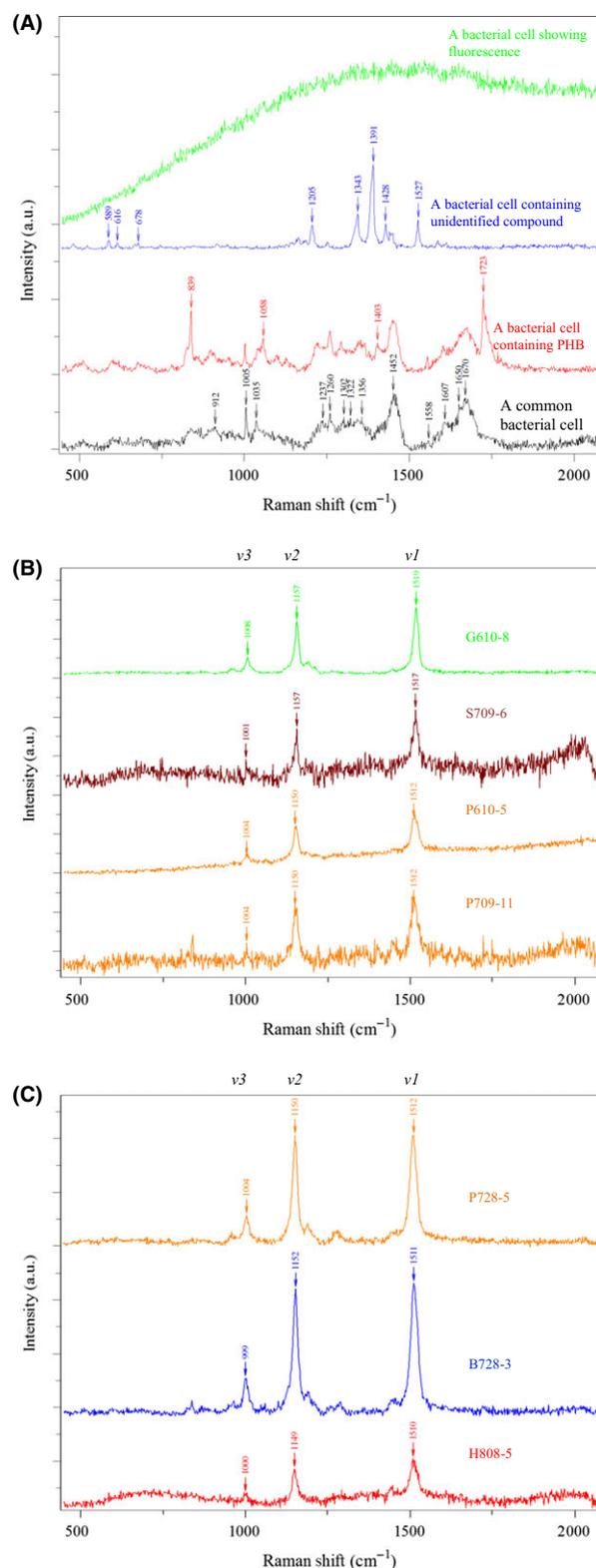


Fig. 3. (A) Single cell Raman spectra of some typical bacteria in a Red Sea sample. (B) Raman spectra of four groups of ejected cells. Each group contains 3–8 cells which had the same carotenoid spectra. (C) Raman spectra of three groups of ejected individual cells. The positions of the carotenoid ν_1 , ν_2 and ν_3 bands are labelled.

Table 1. Summary of RACE-isolated cells and WGA results

Sample name	P728-5	B728-3	H808-5	S709-6	P709-11	G610-8	P610-5
Cells being ejected	1	1	1	3	5	8	8
Identification	<i>Pelomonas</i> spp.	<i>Bradyrhizobium</i> spp.	<i>Halomonas</i> spp.	<i>Shigella</i> spp.	<i>Pelomonas</i> spp.	<i>Cyanobacteria</i> spp.	<i>Pelomonas</i> spp.
Genome coverage (%)	8.18	6.65	4.17	10.74	6.9	8.95	19.29
Contamination	0	0		0	0	0	0
Putative genes for carotenoid biosynthesis	<i>crfI</i>	<i>ispA, crfE, crfI</i>	<i>shc^a</i>	<i>crfI</i>	<i>idi, ispA, crfE, crfI</i>	<i>idi</i>	<i>crfI, crfC, crfJ</i>
Putative genes for CO ₂ fixation	Aerobic-type <i>coxL/coxL</i> ;				CO ₂ concentrating mechanism/ carboxysome shell protein; <i>pepc</i>		CO ₂ -fixation; Calvin-Benson-cycle related gene
Putative genes for haem/ chlorophyll biosynthesis	<i>urod, cobN, chfD, ppox, fech</i>	<i>cobN, haem biosynthesis</i>	<i>cobN</i>	<i>cobN, uros, chfI</i>	<i>cobN, chlI, chfD, ppox, fech</i>		<i>cobN, chfD, chlI, haem oxygenase, fech, ppodg, cpox</i>
Putative genes for fatty acid, alcohol, aldehyde metabolism	acc; Acyl-CoA synthetases, short-chain alcohol dehydrogenases, <i>mcat, Molybdenum cofactor biosynthesis enzyme</i>	Short-chain alcohol dehydrogenases, short-chain dehydrogenases, Acyl-CoA synthetases, <i>fhl</i> subunit 4	Short-chain alcohol dehydrogenases; Short-chain dehydrogenases; Acyl-CoA synthetases; acc; uncharacterized <i>fdh</i>	Acyl-CoA synthetases; <i>hadh; acc;</i> short-chain alcohol dehydrogenases; Predicted acyltransferases; <i>fhl</i> subunit 3 and 4, <i>mmhD</i> subunit; uncharacterized protein for <i>fdh</i> activity	Acyl-CoA synthetases; Alcohol dehydrogenase, class IV; NAD ⁺ ; uncharacterized protein for <i>fdh</i> activity	Alternatively TCA cycle; short-chain alcohol dehydrogenases; Acyl-CoA synthetases; Acyl-CoA	Acyl-CoA synthetases; <i>fhl</i> subunit 3 and 4, <i>mmhD</i> subunit

^aThe substrate of Shc (Squalene-hopene cyclase) is a carotenoid compound.

acc, acetyl-CoA carboxylase; carboxyltransferase component; *acp*, acyl carrier protein (used in FA synthesis); *chfD, chlI*: Mg-chelatase subunit; *cobN*, cobalamin biosynthesis protein CobN and related Mg-chelatases; *coxL/coxL*: homologues, carbon monoxide dehydrogenase, large subunit homologues (used in carbon fixation via reductive acetyl-CoA pathway); *cpox*, coproporphyrinogen III oxidase; *crfC*, hydroxyneurosporene synthase; *crfE*, geranylgeranyl pyrophosphate synthase; *crfI*, phytoene desaturase (dehydrogenase); *crfJ*, CrtJ protein which is involved in spheroidene biosynthesis; *fdh*, formate dehydrogenase; *feh*, Protoheme ferro-lyase (ferrochelatase); *fhl*, formate hydrogenase; *hadh*, 3-hydroxyacyl-CoA dehydrogenase; *idi*, isopentenyl diphosphate isomerase (or IPI, isopentenyl pyrophosphate isomerase); *ispA*, farnesyl diphosphate synthase; *mcat*, malonyl-CoA:acyl carrier protein transacylase; *mmhD*, Na⁺/H⁺ antiporter subunit D; *ppodg*, Porphobilinogen deaminase; *pepc*, Phosphoenolpyruvate carboxylase; *ppox*, Protoporphyrinogen oxidase; *shc*, Squalene-hopene cyclase; *urod*, Uroporphyrinogen-III decarboxylase; *uros*, Uroporphyrinogen-III synthase.

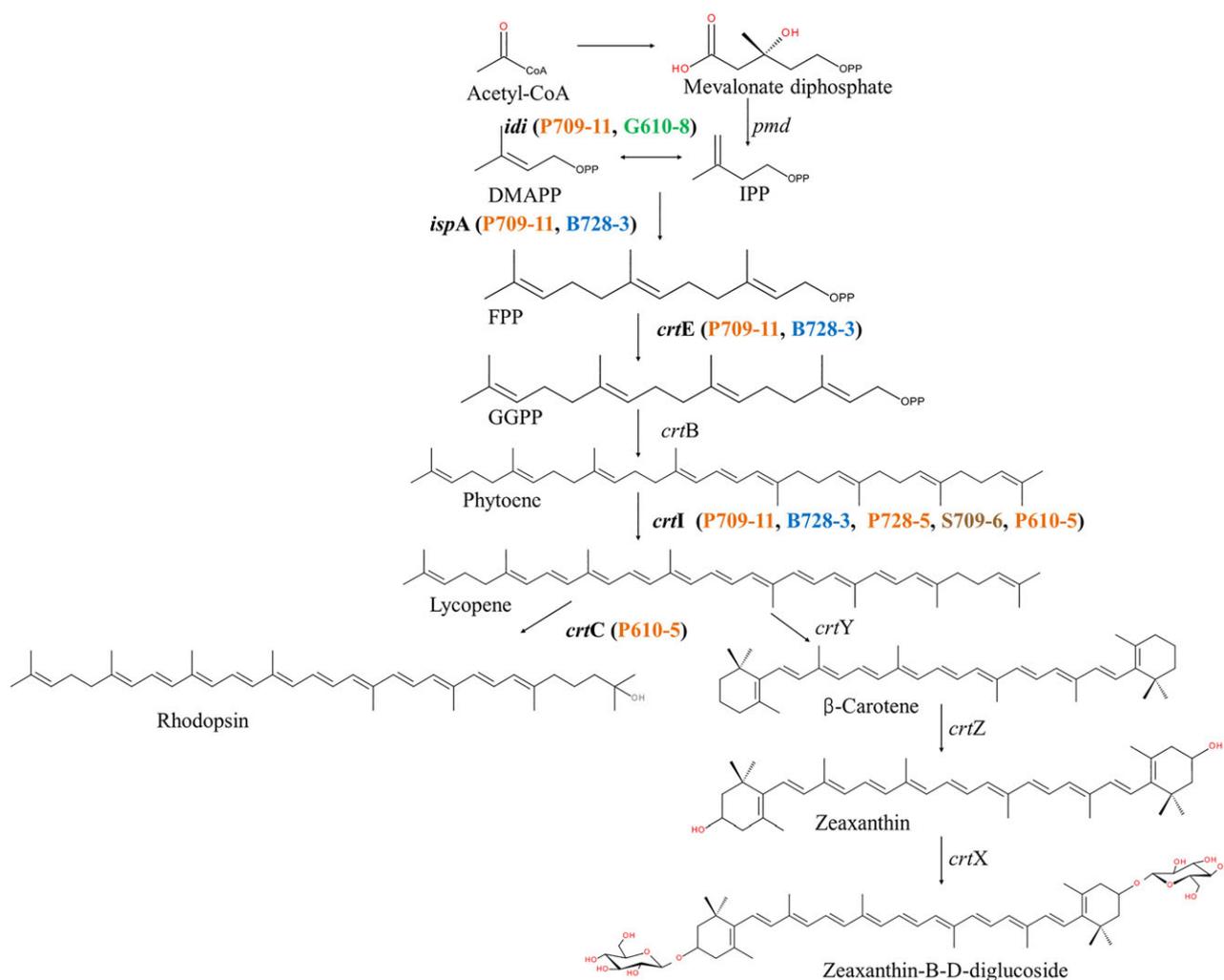


Fig. 4. The identified genes encoding putative proteins in bacterial carotenoid synthesis pathway from Red Sea single cell(s) MDA. The ejected single cell(s) which contain the genes are bracketed and coloured. *idi*: isopentenyl diphosphate isomerase; *ispA*: farnesyl diphosphate synthase; *crtE*: Geranylgeranyl pyrophosphate synthase; *crtB*: phytoene synthase; *crtI*: phytoene desaturase (dehydrogenase); *crtC*: hydroxyneurosporene synthase; *crtD*: methoxyneurosporene dehydrogenase; *crtF*: hydroxyneurosporene-O-methyltransferase; *crtY*: lycopene cyclase; *crtZ*: carotene hydroxylase; *pmd*: phosphomevalonate decarboxylase.

P709-11, S709-6, G610-8 and P610-5). The recovered genes, which are directly involved in carotenoid biosynthesis pathways, are shown in Fig. 4. Specifically, isopentenyl diphosphate isomerase and geranylgeranyl pyrophosphate synthase (*ispA*) are involved in making the colourless substrate farnesyl pyrophosphate, and phytoene dehydrogenase (*crtI*) is responsible for synthesis of the red pigment lycopene (Fig. 4). H808-5 contained two types of putative *shc* (squalene-hopene cyclase) genes whose substrates are carotenoid compounds (Table 1), although no gene directly related to carotenoid synthesis was found in this sample (probably due to low genome coverage). These results validate the RACE method since the sorted cells were expected to contain carotenoids according to their SCRS.

A phylogenetic tree of the isolated samples is shown in Fig. 5. The carotenoid-containing bacteria sorted from the Red Sea sample were broadly distributed over several different groups. According to 16S rRNA sequencing and BLAST (Altschul *et al.*, 1990) against the NCBI bacterial 16S rRNA database, one of the sorted cells, G610-8, is a new cyanobacterial order (GenBank accession no. KU667126), which gives a best hit (84% homology) to an uncultivated *Melainabacteria* spp. (Fig. 5). The 1.2 kb 16S rRNA sequencing has been independently repeated four times by three Sanger sequencing runs and one Illumina MiSeq sequencing run. All sequencing runs gave identical results, which rules out sequencing errors. This microorganism is described as a novel, as yet uncultivated order of cyanobacteria (Soo *et al.*, 2014).

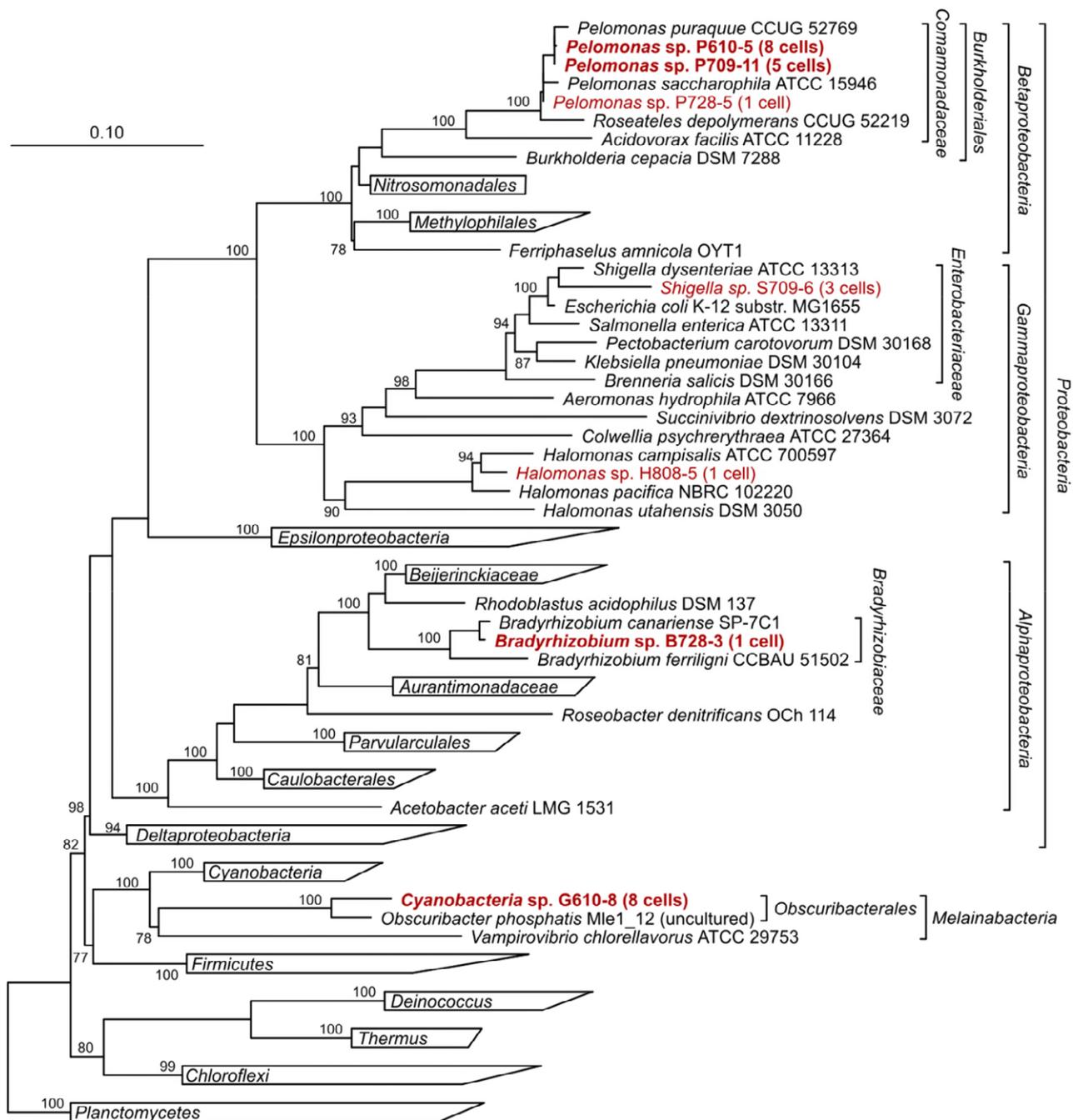


Fig. 5. Neighbor joining tree indicating the phylogenetic relationships of cells sorted by Raman-activated cell ejection and reference genomes. The tree is based on 16S rRNA gene sequences. Bootstrap support values above 75% are indicated at the respective nodes. Leaves referring to 16S rRNA gene sequences obtained from PCR products of MDA treated sorted cells, using universal bacterial primers, are marked in plain red lettering. Leaves referring to 16S rRNA gene sequences obtained from assembled single-cell genome sequencing data of sorted cells are marked in bold red lettering. Black lettering indicates reference sequences.

Discussion

The RACE technology has been applied to a Red Sea water sample to sort seven groups of cells and perform subsequent genome analyses (Table 1) and is a proof of concept for RACE applications in single-cell genomics.

MDA and genome recovery

There are further refinements needed to improve this technique in future studies, such as keeping the sorted cells viable, refining single-cell lysis and genome amplification to increase genome coverage, and reducing

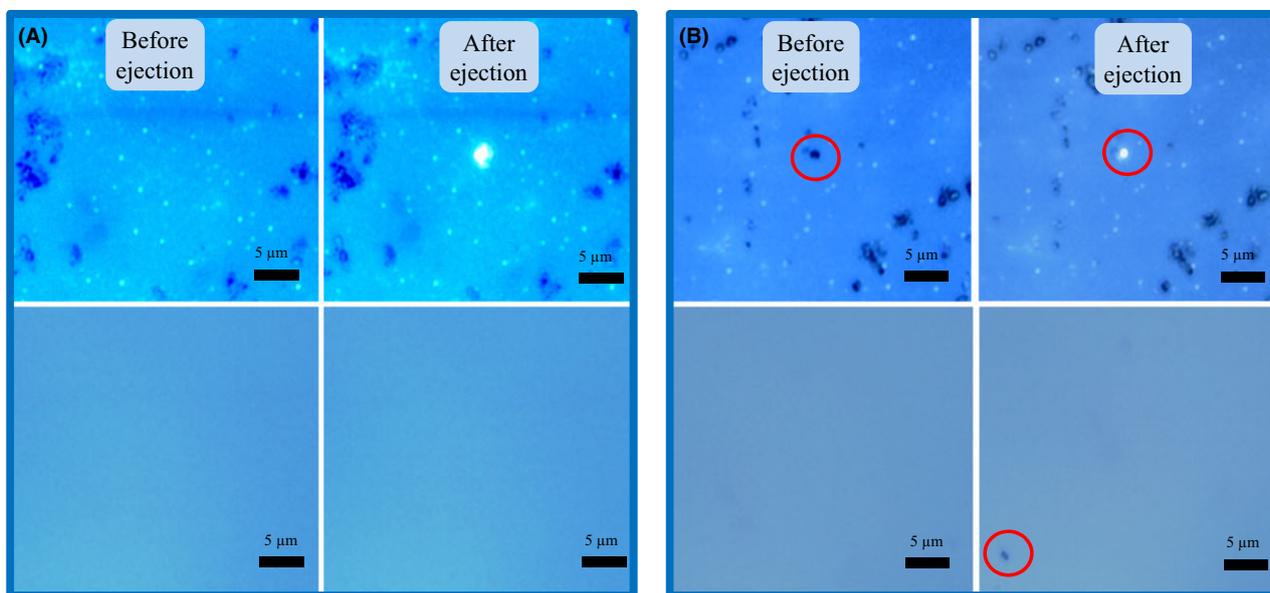


Fig. 6. Microscopy images of the RACE chip (top row) which holds a sample from the Red Sea and the collector (bottom row) before and after applying the pulsed laser. (A) Pulsed laser focused at a position without cells. (B) Pulsed laser focused on a cell.

contamination through better designed RACE chips. The low genome recovery is probably due to the following issues. (i) The cells may not be deposited into the lysis buffer in microwell (Fig. S1), the trajectory of cells can be further improved by adjusting the power of pulsed laser and exposure time. (ii) UV treatment of the Phi29 polymerase and inhibition residual hyperchloride used in the process, resulting in a loss of activity. Highly pure and DNA-free Phi29 polymerase is now available in commercial market. (iii) The cells were dried on the chip for several weeks, and genome DNA may be damaged or degraded during the time. It can be improved by using fresh cells.

Accurate single-cell ejection

The thin coating material was able to absorb the energy of the 532 nm pulsed laser and to eject a single cell sitting on it (Fig. 6 and Fig. S2). Figure 6A and Fig. S2A show that the coating material got ablated after laser ejection, leaving a $\sim 1.5 \mu\text{m}$ mark on the coating slide. No debris could be found on the collection chip after ejection in the blank controls, suggesting that the coating material was completely vapourized. Figure 6B and Fig. S2B indicate that single cells could be accurately isolated and collected by the RACE collection chip.

For single-cell isolation using LIFT, the pulsed laser causes local vapourization of laser absorbing materials in the coating layer of the glass slide, which pushes the cells sitting on it to the collection well (Fig. 6 and Fig. S2). It has been shown that a 532 nm pulsed laser was able to vapourize metallic materials such as Cu and

Ag (Bohandy *et al.*, 1988). It has been previously demonstrated that cells can survive after isolation by LIFT, including bacterial cells (*E. coli*), various mammalian cells and sensitive mouse embryonic stem cells (Pique *et al.*, 1999; Fernandez-Pradas *et al.*, 2004; Ringeisen *et al.*, 2004; Hopp *et al.*, 2005; Kattamis *et al.*, 2007) (Ringeisen *et al.*, 2002).

To achieve RACE, the thin metal coating in the chip should satisfy two criteria: (i) give a minimal Raman background (Fig. 1C), which ensures high-quality SCRS for distinguishing phenotypes; (ii) be able to absorb a pulsed laser and cause *local* vapourization and expulsion of cells without thermal damage. After testing a variety of coating slides, a coating slide provided by Hesen Biotech (Shanghai, China) satisfied the above criteria as it has no Raman signal and absorbs a 532 nm pulsed laser causing ejection of cells.

Discovering new genes using RACE

Three sorted samples (P728-5, P709-11 and P610-5) with similar resonant Raman spectra (Fig. 3) were identified as closely related *Pelomonas* spp. according to phylogenetic analysis (Fig. S6A). Sequence identities in overlapping genome regions even indicate that these genomes may belong to the same species, but due to the low genome coverage resulting in only a few overlaps between the genomes, this could not be verified with confidence. *Pelomonas* is an understudied genus, and there are only a few reports concerning their metabolism, one suggesting that *Pelomonas* spp. could grow chemolithoautotrophically with hydrogen, and indicating the presence of CO_2 fixation

genes (Gomila *et al.*, 2007). The sequences obtained from RACE-based genomes of three *Pelomonas* spp. all included carotenoid and haem/chlorophyll biosynthesis and CO₂ fixation genes, suggesting that these cells might be able to perform photosynthesis in their natural habitat, the Red Sea. Genes identified that are involved in CO₂ fixation include phosphoenolpyruvate carboxylase (from P709-11) which converts bicarbonate to oxaloacetate; a gene involved in making carboxysomes (for CO₂ concentrating (from P709-11); and a CO dehydrogenase that is involved in carbon fixation *via* the reductive acetyl-CoA pathway (from P728-5). It is possible that these cells are chemolithotrophs/oligotrophs and use these functional genes to survive in an oligotrophic system such as Red Sea.

A few genes encoding haem/chlorophyll biosynthesis enzymes were also recovered in the three *Pelomonas* spp., as well as in the *Bradyrhizobium* spp. B728-3, *Halomonas* spp. H808-5 and *Shigella* spp. S709-6. Those genes included magnesium chelatase subunits, ferrochelatase and several enzymes upstream involved in protoporphyrin IX synthesis (Table 1 and summary of genome sequences). After performing a best match search against the NCBI GenBank database (Altschul *et al.*, 1990), putative novel genes were found in the sorted samples. These genes have homology to well-characterized genes but contain novel sequences. For example: a novel gene fragment (GenBank accession no. KX246394) encoding geranylgeranyl pyrophosphate synthase (CrtE) was found in *Bradyrhizobium* spp. B728-3 (Fig. S7A); and a gene (GenBank accession no. KX246395) encoding CobN/magnesium chelatase from *Pelomonas* spp. P728-5 is 87% homologous to the cobalamin gene *cobN* in *Bradyrhizobium* sp. S23321 DNA and 71% homologous to a gene encoding a CobN/magnesium chelatase in *Rhodobacter sphaeroides* (Fig. S7B). Alignment of two putative *shc* genes found in H808-5 genomic sequence (Fig. S7C and D, GenBank accession no. KX246396 and KX246397) indicates that they are novel. SHC encoded by *shc* can catalyse very complex reactions, and any chemical mimicry of these reactions has so far been unsuccessful (Siedenburg and Jendrossek, 2011). Using single-cell genomics, a link between genes and the identities of their host cells can be established, which may reveal the function of these microorganisms in their natural environment. This approach illustrates how genomic studies at the single-cell level can provide information on unculturable and so far unknown microorganisms.

RACE is able to distinguish cells with different Raman spectra and fluorescence

Raman micro-spectroscopy is not only able to distinguish cells with different Raman spectra but also can be used

to detect fluorescence at the single-cell level which usually produces a large broad 'bump' spectrum dwarfing the Raman signal of SCRS (Fig. 3A). Fluorescence in some cells can be photo-bleached by extending laser exposure time (Huang *et al.*, 2007a,b). Since SCRS is independent of laser excitation, fluorescence interference in SCRS can be overcome by changing the incident lasers. SCRS are usually more informative than fluorescence. A RACE system able to distinguish and sort cells based on SCRS could, therefore, be more useful for the study of bacterial phenotypes.

Conclusion

To our knowledge, this is the first report of the use of Raman-activated cell sorting coupled to single-cell genomics. Single-cell genomics based on Raman sorting can not only sort cells with specific compounds (e.g. carotenoids in this case) but also isolate cells with carbon, nitrogen and general metabolic activity when it is coupled with stable isotope probing (Radajewski *et al.*, 2000), which would help link the specific metabolism of single cells (e.g. carbon, nitrogen substrate metabolism or general metabolic activity) (Huang *et al.*, 2004, 2007a,b; Li *et al.*, 2012a,b,c; Wang *et al.*, 2013; Berry *et al.*, 2015) and assist in defining the ecological functions of uncultivated bacteria in the environment.

Experimental procedures

Chemicals, microorganisms and growth conditions

All chemicals used in this study were purchased from Sigma-Aldrich (Dorset, UK) unless otherwise stated.

Escherichia coli JM109 (Promega, Southampton, UK) containing a p18GFP plasmid was incubated at 37°C in LB broth supplemented with 100 µg ml⁻¹ ampicillin.

Seawater samples

The seawater sample was collected from the pier of the Inter-University Institute for Marine Sciences (IUI) in Eilat, Israel, on May 2015. The pier is 50 m long and the water depth is 5 m. The samples were concentrated 20 times using a Centricon Plus-70 Ultracel PL-100 (Merck Millipore, Tullagreen Carrigtwohill, Ireland) and then sent to the laboratory in Oxford on ice.

Confocal Raman micro-spectroscopy and spectral processing

Cells in concentrated seawater samples were washed before being analysed by Raman micro-spectroscopy to remove salts that interfere with cell observation. Each

cellular suspension (1 μl) was mounted in the designed mini-wells of the RACE chip and allowed to air dry before Raman analysis (Fig. S1). The Raman spectra were acquired using a confocal Raman microscope (LabRAM HR Evolution; HORIBA Scientific, London, UK) equipped with an integrated microscope (BX41; Olympus, Essex, UK). A 50 \times long working distance objective (MPLFLN, NA 0.8; Olympus) was used to observe and acquire Raman signals from single cells and the overall magnification is 500. The laser beam position was calibrated and marked by the software Labspec6 (HORIBA Scientific). Cells were visualized by an integrated colour camera and a motorized XYZ stage (0.1 μm step). The Raman scattering was excited with a 532 nm Nd:YAG laser (Ventus, Laser Quantum, Manchester, UK). Raman measurements with grating 600 l mm^{-1} resulted in a spectral resolution of $\sim 1 \text{ cm}^{-1}$ with 1019 data points. The laser power on a single cell was about 0.5 mW. The detector was a -70°C air-cooled charge coupled device detector (Andor, Belfast, UK). LabSpec6 software was also used to control the Raman system and acquire the Raman spectra. Acquisition times for Raman spectra were 5 s for single-cell measurements. The spatial location of each measured cell from the Red Sea sample were recorded in Labspec6 and used to identify cells for subsequent single-cell ejection.

Integration of single-cell ejection into a Raman microscope

A 532 nm pulsed laser (ALPHALAS GmbH, Goettingen, Germany) was integrated into a confocal Raman microscope, sharing the same continuous laser (532 nm) light path and objective lens for SCRS measurement without changing mirrors and filters (Fig. 1).

The peak power was 1.1 kW, about two magnitudes lower than the traditional UV microdissection lasers (Table S1). The power of this 532 nm pulsed laser is maximally 3.4 mJ cm^{-2} (Table S1), which is much lower than the maximum resistible power (400 mJ cm^{-2}) of the mirror and edge filters in the Raman microscope. An exposure test showed that the 532 nm pulsed laser did indeed cause no damage to the mirror and edge filters in this LabRAM HR Evolution Raman microscope (data not shown).

Sterilization condition and reagents for single-cell genomics

DNA contamination is a major issue in single-cell genomics, due to the very small amount of template DNA derived from a single cell which has to be amplified by WGA. Standard sterilization practices are, therefore, not sufficient for single-cell WGA and a modified WGA method had to be developed (Rinke *et al.*, 2013, 2014) to minimize the influence of contaminant DNA. All the

equipment and working surfaces were wiped with 10% (v/v) domestic bleach (Domestos, Surrey, UK). All the consumables were autoclaved and exposed to a UV light bulb (254 nm) inside a UV sterilized Laminar Flow cabinet for 1 h. All reagents were carefully sterilized, and the detailed process is described in Supplementary Information.

Raman-activated single-cell ejection (RACE)

Supplementary information provides the details of reagent preparation and procedures. Briefly, 1 μl of the lysis buffer and 1 μl TE buffer were added into each well on the collection chip, which was then attached to the RACE chip. The enclosed chips were moved to the stage of the Raman microscope with the RACE chip facing down. The target cells were located by their coordinates and confirmed by comparing the bright-field image taken during Raman spectra acquisition. The laser spot was manually aligned to the coating layer under the target cell. The target cells were ejected by evaporation of the coating layer upon application of a <1 s exposure of the 532 nm pulsed laser, and the cells were harvested in the collector. For a 12-well RACE chip, at least two control wells were also set up: one negative control, which remained cell-free, and one positive control, to which cell suspensions were added before WGA (see Data S1).

Whole-genome amplification on-chip

The enclosed chips were carefully moved to a laminar flow chamber and the two chips were separated. A sterile coverslip was then placed on top of the collector chip. Wells were separated and independent with interference, and each well is a closed space. Cell lysis was carried out with lysis buffer (Qiagen, Manchester, UK) pre-added in the collection chip (see Data S1). Three freeze-thaw cycles were performed to enhance cell lysis. Subsequently, the chip was heated at 65°C for 10 min in a thermocycler (C1000; Bio-Rad, Hemel Hempstead, UK) to ensure a complete lysis of cells. After adding 1 μl of Stop solution to neutralize the lysis buffer, 12 μl of reaction master mix (containing phi29 polymerase from Epicenter, Cambridge, Camlab, UK) was added to each well (Data S1). The collector chip was then covered by a coverslip and kept in the thermocycler at 30°C with the hot lid activated and set at 70°C . After incubation for 8 h, the Phi29 DNA polymerase was deactivated by heating to 65°C for 10 min. The MDA product was then transferred into sterilized 200 μl PCR tubes for storage.

Library preparation and sequencing

Amplified DNA from single cells and cell consortia was quantified using a Qubit 1.0 fluorometer and a dsDNA

HR assay kit (Life Technologies, Darmstadt, Germany). Libraries were constructed using a Nextera XT DNA Library Preparation Kit (Illumina, Cambridge, UK) and 1 ng of amplified input DNA, as per the manufacturer's instructions. Sequencing was done on an Illumina MiSeq machine with paired end settings and 301 cycles per read.

Sequence processing and assembly

The raw sequences were subjected to adapter clipping and quality trimming using Trimmomatic (Bolger *et al.*, 2014) with the following arguments: 'LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:60'. Residual contaminant PhiX reads were removed from the datasets using FastQ Screen (<http://www.bioinformatics.babraham.ac.uk/>). In order to minimize subjective bias, no other potential contaminants from *E. coli* and human DNA were filtered from the datasets. To reduce MDA-induced bias of overrepresented genome regions, the standard three-step digital normalization protocol of the khmer suite was employed (Crusoe *et al.*, 2015). The normalized reads were then assembled using Spades v3.6 (Bankevich *et al.*, 2012).

Sequence analyses

For each assembly, the genome completeness and degree of contamination was estimated using CheckM (Parks *et al.*, 2015). Prokka v1.10 was used for gene prediction and annotation. Universal marker gene products were extracted using fetchMG (<http://www.bork.embl.de/software/mOTU/fetchMG.html>), compared against the NCBI nr database using BLAST and phylogenetically classified using MEGAN5 (Huson and Weber, 2013). For 16S rRNA gene sequences extracted from assembled genomes or generated from screening PCRs, the rough taxonomical placement was determined based on BLAST searches against the NCBI bacterial 16S rRNA database, and suitable reference sequences originating from described type strains were selected accordingly. All 16S rRNA sequences were aligned against the SILVA database using Sina (Pruesse *et al.*, 2012) and incorporated into phylogenetic trees using Arb v5.5 (Ludwig *et al.*, 2004). Only sequences longer than 1200 bp were used for calculation of the basic tree backbones, while sequences shorter than 1200 bp were subsequently added using the parsimony function of Arb.

Conflict of interest

The authors declare no conflict of interest.

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410.
- Amann, R.L., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and in-situ detection of individual microbial-cells without cultivation. *Microbiol Rev* **59**: 143–169.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**: 455–477.
- Berry, D., Mader, E., Lee, T.K., Woecklen, D., Wang, Y., Zhu, D., *et al.* (2015) Tracking heavy water (D₂O) incorporation for identifying and sorting active microbial cells. *Proc Natl Acad Sci USA* **112**: E194–E203.
- Blanco, L., Bernad, A., Lazaro, J.M., Martin, G., Garmendia, C., and Salas, M. (1989) Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. DNA replication. *J Biol Chem* **264**: 8935–8940.
- Bohandy, J., Kim, B.F., Adrian, F.J., and Jette, A.N. (1988) Metal deposition at 532 nm using a laser transfer technique. *J Appl Phys* **63**: 1158–1162.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.
- Bryant, D.A., and Frigaard, N.-U. (2006) Prokaryotic photosynthesis and phototrophy illuminated. *Trends Microbiol* **14**: 488–496.
- Chen, J., and Griffiths, M.W. (1998) PCR differentiation of *Escherichia coli* from other Gram-negative bacteria using primers derived from the nucleotide sequences flanking the gene encoding the universal stress protein. *Lett Appl Microbiol* **27**: 369–371.
- Crusoe, M.R., Alameddine, H.F., Awad, S., Boucher, E., Caldwell, A., Cartwright, R., *et al.* (2015) The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Research* **4**: 900–900.
- Daniel, R. (2005) The metagenomics of soil. *Nat Rev Microbiol* **3**: 470–478.
- Fernandez-Pradas, J.M., Colina, M., Serra, P., Dominguez, J., and Morenza, J.L. (2004) Laser-induced forward transfer of biomolecules. *Thin Solid Films* **453**: 27–30.
- Garcia-Asua, G., Lang, H.P., Cogdell, R.J., and Hunter, C.N. (1998) Carotenoid diversity: a modular role for the phytoene desaturase step. *Trends Plant Sci* **3**: 445–449.
- Gomila, M., Bowien, B., Falsen, E., Moore, E.R.B., and Lalucat, J. (2007) Description of *Pelomonas aquatica* sp nov and *Pelomonas puraquae* sp nov., isolated from industrial and haemodialysis water. *Int J Syst Evol Microbiol* **57**: 2629–2635.
- Gonzalez, J.M., Pinhassi, J., Fernandez-Gomez, B., Coll-Llado, M., Gonzalez-Velazquez, M., Puigbo, P., *et al.* (2011) Genomics of the proteorhodopsin-containing marine flavobacterium *Dokdonia* sp strain MED134. *Appl Environ Microbiol* **77**: 8676–8686.
- Hedlund, B.P., Dodsworth, J.A., Murugapiran, S.K., Rinke, C., and Woyke, T. (2014) Impact of single-cell genomics and metagenomics on the emerging view of extremophile “microbial dark matter”. *Extremophiles* **18**: 865–875.

- Hopp, B., Smausz, T., Kresz, N., Barna, N., Bor, Z., Kolozsvari, L., *et al.* (2005) Survival and proliferative ability of various living cell types after laser-induced forward transfer. *Tissue Eng* **11**: 1817–1823.
- Huang, W.E., Griffiths, R.I., Thompson, I.P., Bailey, M.J., and Whiteley, A.S. (2004) Raman microscopic analysis of single microbial cells. *Anal Chem* **76**: 4452–4458.
- Huang, W.E., Bailey, M.J., Thompson, I.P., Whiteley, A.S., and Spiers, A.J. (2007a) Single-cell Raman spectral profiles of *Pseudomonas fluorescens* SBW25 reflects in vitro and in planta metabolic history. *Microb Ecol* **53**: 414–425.
- Huang, W.E., Stoecker, K., Griffiths, R., Newbold, L., Daims, H., Whiteley, A.S., and Wagner, M. (2007b) Raman-FISH: combining stable-isotope Raman spectroscopy and fluorescence in situ hybridization for the single cell analysis of identity and function. *Environ Microbiol* **9**: 1878–1889.
- Huang, W.E., Ferguson, A., Singer, A., Lawson, K., Thompson, I.P., Kalin, R.M., *et al.* (2009a) Resolving genetic functions within microbial populations: in situ analyses using rRNA and mRNA stable isotope probing coupled with single-cell Raman-fluorescence in situ hybridization. *Appl Environ Microbiol* **75**: 234–241.
- Huang, W.E., Ward, A.D., and Whiteley, A.S. (2009b) Raman tweezers sorting of single microbial cells. *Environ Microbiol Rep* **1**: 44–49.
- Huang, W.E., Song, Y., and Xu, J. (2015) Single cell biotechnology to shed a light on biological 'dark matter' in nature. *Microb Biotechnol* **8**: 15–16.
- Huson, D.H. and Weber, N. (2013) Microbial community analysis using MEGAN. In *Microbial Metagenomics, Metatranscriptomics, and Metaproteomics*. DeLong, E.F. (ed.). Elsevier Inc. The Netherlands: Chapter 21, pp. 465–485.
- Kattamis, N.T., Purnick, P.E., Weiss, R., and Arnold, C.B. (2007) Thick film laser induced forward transfer for deposition of thermally and mechanically sensitive materials. *Appl Phys Lett* **91**: 171120.
- Kochenderfer, G.G., Lin, S.W., Sakmar, T.P., and Mathies, R.A. (1999) How color visual pigments are tuned. *Trends Biochem Sci* **24**: 300–305.
- Krebs, R.A., Dunmire, D., Partha, R., and Braiman, M.S. (2003) Resonance Raman characterization of proteorhodopsin's chromophore environment. *Journal of Physical Chemistry B* **107**: 7877–7883.
- Lasken, R.S. (2012) Genomic sequencing of uncultured microorganisms from single cells. *Nat Rev Microbiol* **10**: 631–640.
- Li, M., Xu, J., Romero-Gonzalez, M., Banwart, S.A., and Huang, W.E. (2012a) Single cell Raman spectroscopy for cell sorting and imaging. *Curr Opin Biotechnol* **23**: 56–63.
- Li, M., Canniffe, D.P., Jackson, P.J., Davison, P.A., FitzGerald, S., Dickman, M.J., *et al.* (2012b) Rapid resonance Raman microspectroscopy to probe carbon dioxide fixation by single cells in microbial communities. *ISME J* **6**: 875–885.
- Li, M., Ashok, P.C., Dholakia, K., and Huang, W.E. (2012c) Raman-activated cell counting for profiling carbon dioxide fixing microorganisms. *J Phys Chem A* **116**: 6560–6563.
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Kumar, Y., *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.
- Majed, N., and Gu, A.Z. (2010) Application of Raman microscopy for simultaneous and quantitative evaluation of multiple intracellular polymers dynamics functionally relevant to enhanced biological phosphorus removal processes. *Environ Sci Technol* **44**: 8601–8608.
- Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043–1055.
- Paterson, E., Hall, J.M., Rattray, E.A.S., Griffiths, B.S., Ritz, K., and Killham, K. (1997) Effect of elevated CO₂ on rhizosphere carbon flow and soil microbial processes. *Glob Change Biol* **3**: 363–377.
- Pique, A., Chrisey, D.B., Auyeung, R.C.Y., Fitz-Gerald, J., Wu, H.D., McGill, R.A., *et al.* (1999) A novel laser transfer process for direct writing of electronic and sensor materials. *Applied Physics a-Materials Science & Processing* **69**: S279–S284.
- Pruesse, E., Peplies, J., and Gloeckner, F.O. (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823–1829.
- Radajewski, S., Ineson, P., Parekh, N.R., and Murrell, J.C. (2000) Stable-isotope probing as a tool in microbial ecology. *Nature* **403**: 646–649.
- Ringeisen, B.R., Chrisey, D.B., Pique, A., Young, H.D., Modi, R., Bucaro, M., *et al.* (2002) Generation of mesoscopic patterns of viable *Escherichia coli* by ambient laser transfer. *Biomaterials* **23**: 161–166.
- Ringeisen, B.R., Kim, H., Barron, J.A., Krizman, D.B., Chrisey, D.B., Jackman, S., *et al.* (2004) Laser printing of pluripotent embryonal carcinoma cells. *Tissue Eng* **10**: 483–491.
- Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.F., *et al.* (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431–437.
- Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., *et al.* (2014) Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat Protoc* **9**: 1038–1048.
- Robert, B. (2009) Resonance Raman spectroscopy. *Photosynth Res* **101**: 147–155.
- Schleifer, K.H. (2004) Microbial diversity: facts, problems and prospects. *Syst Appl Microbiol* **27**: 3–9.
- Siedenburg, G., and Jendrossek, D. (2011) Squalenohopene cyclases. *Appl Environ Microbiol* **77**: 3905–3915.
- Soo, R.M., Skennerton, C.T., Sekiguchi, Y., Imelfort, M., Paech, S.J., Dennis, P.G., *et al.* (2014) An expanded genomic representation of the phylum cyanobacteria. *Genome Biology and Evolution* **6**: 1031–1045.
- Swan, B.K., Martinez-Garcia, M., Preston, C.M., Sczyrba, A., Woyke, T., Lamy, D., *et al.* (2011) Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* **333**: 1296–1300.
- Takaichi, S. (2008) Distribution and biosynthesis carotenoids. In *The Purple Phototrophic Bacteria*. Hunter, C.N. (ed.). The Netherlands: Springer Science, pp. 97–117.
- Venter, J.C. (2003) Unleashing the power of genomics: understanding the environment and biological diversity. *Scientist* **17**: 8–8.

- Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.
- Wang, Y., Ji, Y., Wharfe, E.S., Meadows, R.S., March, P., Goodacre, R., *et al.* (2013) Raman activated cell ejection for isolation of single cells. *Anal Chem* **85**: 10697–10701.
- Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998) Prokaryotes: the unseen majority. *Proc Natl Acad Sci USA* **95**: 6578–6583.
- Zhang, Q., Zhang, P., Gou, H., Mou, C., Huang, W.E., Yang, M., *et al.* (2015a) Towards high-throughput microfluidic Raman-activated cell sorting. *Analyst* **140**: 6163–6174.
- Zhang, P., Ren, L., Zhang, X., Shan, Y., Wang, Y., Ji, Y., *et al.* (2015b) Raman-activated cell sorting based on dielectrophoretic single-cell trap and release. *Anal Chem* **87**: 2282–2289.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Data S1. Materials and methods.

Fig. S1. Illustration of the RACE chip design.

Fig. S2. Isolation of a single *E. coli* cell using the RACE chip.

Fig. S3. (A) Agarose gel image of multiple displacement amplifications (MDAs) showing high-molecular-weight DNA.

Fig. S4. PCA axis 1 loadings.

Figure S5. (A) Microscopy image of cells in Fig. 2A and B.

Fig. S6. (A) Phylogenetic tree of three *Pelomonas* spp. P728-5, P709-11 and P610-5 isolated in this study.

Fig. S7. (A) The partial gene encoding CrtE from *Bradyrhizobium* spp. B728-3 (1 cell) is novel and related to *crtE* genes from Cyanobacteria *Synechocystis* spp. and *Gloeobacter* spp. according to BLAST analysis.

Table S1. Pulsed laser power comparison for cell isolation.

Table S2. Primers used in this study.

Table S3. Genome coverage and contamination assessment.

Table S4. Summary of genome sequences.