Ecology of Marine Bacteroidetes: A Genomics Approach

Beatriz Fernández Gómez

Marzo 2012

UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA

Institut de Ciències del Mar

D. Juan Luís Gómez Pinchetti SECRETARIO DEL DEPARTAMENTO DE BIOLOGÍA DE LA UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA,

CERTIFICA,

Que el Consejo de Doctores del Departamento en su sesión extraordinaria tomó el acuerdo de dar el consentimiento para su tramitación a la tesis doctoral titulada *"Ecology of Marine Bacteroidetes: A genomics approach"* presentada por la doctoranda Dª Beatriz Fernández Gómez y dirigida por los Doctores Carlos Pedrós-Alió y José M. González Hernández.

Y para que así conste, y a efectos de lo previsto en el Artº 73.2 del Reglamento de Estudios de Doctorado de esta Universidad, firmo la presente en Las Palmas de Gran Canaria, a .... de ..... de 2012

UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA
Departamento de Biología

Institut de Ciències del Mar
CSIC

Ecology of Marine Bacteroidetes: A Genomics Approach
(Ecología de los Bacteroidetes Marinos: Una aproximación genómica)

Beatriz Fernández Gómez

Tesis Doctoral presentada por Dª Beatriz Fernández Gómez para obtener el grado de Doctora en Oceanografía por la Universidad de Las Palmas de Gran Canaria.

Directores: Dr. Carlos Pedrós-Alió y Dr. José M. González

Universidad de Las Palmas de Gran Canaria, Departamento de Biología
Institut de Ciències del Mar (ICM-CSIC)

DOCTORADO EN OCEANOGRAFÍA
Bienio 2006-2008. Con Mención de Calidad de la ANECA

La Doctoranda	El Director	El Co-director
Beatriz Fernández Gómez	Carlos Pedrós-Alió	José M. González

En Barcelona, a de de 2012

"Somewhere, something incredible
is waiting to be known"
Carl Sagan

A mis padres y Victor
A mi abuela,
A Pedro

Cover: **Water, origin of life.** This is an homage to the life-creating power of water. Life begins at the center, where the first cells and the microscopic universe are represented (Diego Rivera, 1951).
Back-cover: **The Biodiversity Mandala**. Representation of all living creatures of our Planet in proportion of the number of currently described species. Created by Stéphan Daigle by request of National Geographic Magazine (1999).

Portada: **El agua, origen de la vida**. Homenaje al poder creador del agua. La vida comienza en el centro, donde están representadas las primeras células y el universo microscópico (Diego Rivera, 1951).
Contraportada: **El Mandala de la Biodiversidad**. Representación de todas las criaturas vivas en nuestro Planeta en proporción al número de especies actualmente descritas. Creado por Stéphan Daigle a petición de National Geographic Magazine (1999).

AGRADECIMIENTOS/ACKNOWLEDGMENTS

Me gustaría empezar esta tesis dando las GRACIAS a todas las personas que la han hecho posible y a todas aquellas que me han apoyado y dado ánimos en el transcurso de todo el doctorado. ¡Qué guay! Unas cuantas páginas para poder expresarme a mis anchas, sin tener que preocuparme por el lenguaje científicamente correcto! Uff, espero que no se me olvide nadie!

Voy a empezar nombrando a mis dos jefes, la cara y la cruz de la misma moneda, dos polos opuestos, pero que sin embargo se acaban atrayendo para formar un ente excepcional. Carles, en el mismo instante en que Ricard me habló de ti pensé.. wow.. el pupilo de Brock! El creador de ese libro gruesísimo que me acompañó durante toda la carrera, que tenía subrayadísimo y lleno de notas y que además me encantaba, entonces me dije… ese tiene que ser mi jefe!. Qué te puedo decir boss, mil gracias por estos "ajium ajium..." años de doctorado, en los cuales he disfrutado a tope cada curso, cada congreso, la campaña al Ártico.., tantos sitios maravillosos que, como me las mando a mi misma, me han llenado la pared de postales!! Gracias por transmitirme ese entusiasmo por la ciencia y su divulgación y ese interés por lo desconocido y lo místico.
Jose, sin ti creo que hubiera fallado en la primera etapa de mi doctorado. Gracias por enseñarme y ayudarme tanto con la bioinformática y con los cultivos, por "Skypearme" para animarme y preocuparte de cómo estaban yendo las cosas. Aunque hemos estado trabajando en la distancia, siempre te has preocupado y me has ayudado en todo lo que he necesitado. Y gracias por la acogida en Tenerife y por la excursión al Teide!

Y ahora, y no por eso menos importante, el resto de personajes maravillosos!

Silvia Acinas, para mi, mi segunda codirectora. Qué puedo decir de ti bella, mi más sincera gratitud por estar siempre ahí, por tener siempre una idea o consejo para todo. Por transmitir ese positivismo, por hacerme reír, por la oportunidad de ir al TARA, en fin... por ser como eres, una tía genial!

Gracias también al resto de jefazos. Cèlia (espero que ahora ya puedas distinguirme de Raquel, jeje), Dolors, Pep, Ramon, Montse S. y Rafel por estar siempre ahí para un consejo, para ofrecer ayuda. A Miquel Alcaráz y Albert Calbet que me abrieron las puertas del ICM cuando todavía era una universitaria en prácticas (nunca pensé que me quedaría tanto!). Disfruté y aprendí mucho con los copepoditos.

Muy especialmente estos agradecimientos van para mis chicas, mis niñas (y algún que otro niño!). Gracias por cada uno de los momentos vividos a vuestro lado, buenos o malos, por reírnos tanto juntas y por apoyarnos las unas a las otras. Mi recuerdo más especial es para mi Lore, mi mayor motivación y confidente. Gracias por darme calma, por escuchar todos mis problemas y frustraciones. Por compartirlo todo conmigo, por hacerme reír, por tus "sipos", "puchas" y "chutas", que ya forman parte de mi vocabulario. Fue un privilegio poder estar mesa con mesa y tenerte siempre a mi ladito. Ay! Cuanto te estoy echando de menos! Te quiero mucho! Raquel, mi Reich, gracias por las excursioncillas y viajes tan guapos que nos hemos marcado juntas, gracias por tu sonrisa constante y contagiosa. Qué iba a hacer yo (y el resto) sin tu perfecta organización y planificacion! Ah, y de tu coordinación en Bollywood (ahora eres todo un referente para mi! jijí). Montse, simplemente gracias por ser como eres, nunca cambies. Me encanta tu espontaneidad, tu risa picaresca y tu arte de hacer que todo tipo de conversación acabe en el mismo tópico de siempre! Ets la millor! Eli la más cuerda y sensata de todas, gracias por emanar tanta serenidad que contraresta mi nerviosismo. Ero, otra nerviosilla, cada sonrisa tuya vale por mil, ves? tanto

que decías y al final el momento llegó y has sido doctora antes que yo! a qué mereció la pena? Gracias por haberte quedado con nosotras un poquito más. Y ahora mi niño, el gran Massimo. Grazie mile por tus divertidas conversaciones, por los cafés matutinos, por tus "coco-colos" a media tarde y por compartir conmigo las frustraciones que da trabajar delante del ordenador (viva el copypaste!), me has hecho reír muchísimo y no sabes cuanto te lo agradezco. Y a Juanlu, porque no se cómo lo haces pero siempre estás (o te veo) de buen humor y transmitiendo buena onda. Y como no, recordar a mi Verito, mi guají asturiana, aixxx cuánto hemos vivido junticas. Fuiste la primera en marchar pero siempre estás conmigo, vuelveeeee!!!!

A mis otras niñas del despacho, Cris, Elena y Ana (fofinhaaa). Gracias por hacerme sentir que estoy en el mejor despacho del instituto, por aguantar mis momentos de locura, mis histerias, mis voces y portazos que a veces desconcentran un poquito, verdad?. Gracias por hacerme reír, por nuestros momentos de cotis, ..., pero no me voy a despedir demasiado porque todavía nos quedan algunos meses juntas eh, todavía muchas risas que compartir y... vuestras defensas!!

A mis ex-compis de despacho, Javi y Clara L., que empezaron esta aventura conmigo, en aquel lindo y soleado despacho (aixx, como molaba eh!). Espero que os vaya muy bien con vuestros respectivos post-docs.

A los "post-docs", por su gran ayuda: Hugo S., Bea D., Isabel F. y Marta S., gracias, habéis sido un gran apoyo y una referencia, sois magníficos! Y a ti Ramiro, boludo, gracias por resolver mis dudas acerca de como hacer arbolitos, por tu ayuda y los buenos momentos en Hawaii y por los whatever-day beer!

Gracias Vane, Irene y Clara C. por vuestra dedicación y por cuanto me hayáis ayudado en el laboratorio y por estar cuando os necesitaba. Y a Mara por su paciencia al contarme las ampollitas para el DOC!

Y qué decir de las nuevas generaciones que han ido apareciendo, a Guillem (SalaTZar!), Ana Mari, Fran, Sarah-Jeanne, por hacerme rejuvenecer y por transmitirme ese entusiasmo y alegría con la que empezáis. Seguid así chicos que este es un camino muy largo y lo importante es disfrutarlo!

A todos aquellos con los que he compartido momentos de histeria, ánimo mútuo, alegría, frustración, etc., por los pasillos del instituto: Clara R., Arancha, Cris R., Lorenzo, Juancho, Claudio, Mireia, Pati, Julia B., Silvia A., Bego, Rodrigo A., Martí, Thomas L., Eli S., Suso, Mariona, Sofia K., Marco T., Marc, Miriam, Meri, Xavi Leal, aiss que todavía se me olvidará gente..

A mis colegas del CEAB, a Emilio Casamayor por sus consejos y recomenaciones, al Barbe (cuántos bio-momentos vividos!), a Toni, "the geek", porque la ciencias no es todo en la vida, a veces hace falta un poco de fiesta, eh?

To my colleagues from MPI, my German family, who made unforgettable my stay in the cold and dark Bremen. Danke schön Michi, Daniel ("el pinche"), Julia, Francesca, Ilaria, Paola, Ivolito, Elmar, Thierry, Renzo and Christian (Thanks to keep the "Genoma Pitufo" and "Bea was here"-post-it still in my desk!).

Je dois aussi remercier Thomas Pommier, bien qu´il ne soit plus chez ICM, merci pour m´avoir aidé au tout début de mon doctorat. Je veux te remercier pour tes conseils qui, même si tu sais que je ne les prenais pas trop bien, étaient pour mon bien et pour me permettre d´arriver là où je suis aujourd´hui. Je voudrais aussi te remercier pour ces moments de franches rigolades que

nous avons eu ensemble. Merci beaucoup chico!

To F.O Glöckner and Marga Schüler for the chance of working in their microbial group. To Jarone Pinhassi and Laura G-C for their sympathy and help with the cultures. To Connie Lovejoy for the wonderful experience in the Arctic. To Kriss Rokkan for giving me the opportunity to be at 78ºN!. To Dave Karl, Matt Church, Greig Steward and Mike Rappé for the magnificent "Microbial Oceanography Course" in Hawaii (it was so great!), and all the people who made possible my attendance to courses and congresses and all the nice people I met there.

A toda mi gente, que ha sufrido tanto como yo mi doctorado, por aguantar mis batallitas de científica loca (o profesora bacteria) y por hacer que me olvidara por un momento de la ciencia y disfrutara de la vida: Estrella, mi Star, linda, gracias por tus ánimos, por nuestras fiestas y viajes juntas y porque aún estando lejos siempre has estado ahí con tu risa, jo jo jo, te quiero un montón nena (Asere, esa mecánica de doctora!). Dani, gracias por escucharme aunque nunca te enteraras de exactamente con qué bacterias trabajo ni qué hacen. To Souzana for being a magnificent person and take care of me, thank you so so much (you can't believe it's done, aren't you?).

Y no me puedo dejar a Xavi de seguridad (que sino se me enfada!) Con la de conversaciones que hemos tenido cuando salía a horas intempestivas del instituto... pues para que veas que me he acordado de ti y espero poder traerte muchas más arenas de los lugares a los que vaya en el futuro.

Y en el último momento tengo que añadir a estos agradecimientos a Aridane, ya que sin su ayuda nunca hubiera sido capaz de tramitar el depósito de esta tesis, la cual habría naufragado en su camino hasta la ULPGC.

A toda mi familia. A mis padres por su paciencia, por su apoyo incondicional a lo largo de todo este tiempo, por sus ánimos y por hacerme sentir que realmente lo podía conseguir. A mi abuela, por estar tan orgullosa y presumir tanto ante todo el pueblo de nieta científica y viajera (aunque realmente no tiene mucha idea, ella presume!) Y a todo el resto de primos y tíos que se han preocupado por mi trabajo. Por fin existe una respuesta para vuestra eterna pregunta de.. pero tú acabarás ya o qué? Siiii, acabo yaaaa!!

Mi último y más especial agradecimiento es para Pedro ("The physicist"). Gracias por tu apoyo incondicional, por no dejarme caer en los momentos de crisis, por intentar resolver todas mis dudas (a veces pregunto demasiando, verdad?), por animarme con mis experimentos y darme siempre nuevas ideas, por tu ayuda con el ODV (maestro!), por todo lo que nos hemos reído (Bazinga!), por los viajes que nos hemos pegado (para cuándo el próximo?). Acaba una parte de la aventura pero lo bueno está por llegar, prepárate.. it's gonna be LE-GEN-DA-RY! ☺

GRACIAS, GRÀCIES, THANK YOU, DANKE, MERCI, GRAZIE, OBRIGADA!!

"La gratitud en silencio no sirve a nadie"

Pray that the road is long.
May there be many a summer morning, when,
with such pleasure, with such joy
you enter ports seen for the first time.

(Ithaca, K. Kavafis)

RESUMEN

El filo de los Bacteroidetes es uno de los grupos más abundantes de bacterias en los océanos. Representan una fracción significativa del bacterioplancton marino, especialmente en las zonas costeras, aunque están globalmente distribuidos. En general los Bacteroidetes son muy eficientes en la degradación de la materia orgánica particulada, y prefirien sustancias poliméricas en vez de sustancias disueltas. Contrariamente a la mayoría de grupos de bacterias marinas, muchos representantes poseen motilidad por reptación y tienden a ser más abundantes adheridos a partículas que en vida libre en la columna de agua. Por lo tanto, probablemente presenten una estrategia diferente a la de las proteobacterias o cianobacterias y, en consecuencia, ocupen nichos diferentes. En conjunto, los Bacteroidetes juegan un papel único e importante en el ciclo del carbono marino. En esta tesis se han utilizado diversos enfoques para comprender la ecología de algunos miembros de este grupo de bacterias.

En primer lugar se llevó a cabo el análisis completo del genoma de un representante de los Bacteroidetes (*Polaribacter* sp. MED152), del que se pudo deducir una estrategia de supervivencia diferente de las de otros grupos. Posteriormente se realizó un análisis comparativo *in silico* (utilizando herramientas bioinformáticas) entre cuatro Bacteroidetes marinos cuyos genomas han sido completamente secuenciados. Encontramos que estas bacterias tenían un gran número de genes implicados en la degradación de polímeros y adherencia a superficies. Un promedio de 1,5% de la totalidad de estos genomas correspondía a genes adquiridos por medio de transferencia horizontal de genes, lo que proporcionaba nuevas adaptaciones y capacidades metabólicas. Esto llevó al análisis de islas genómicas en bacterias marinas, tratando de discernir posibles patrones característicos en los Bacteroidetes. Algunas de las bacterias analizadas tenían el gen de la proteorodopsina (PR), lo que les permite obtener energía extra en la presencia de luz. Dado que el uso de la luz por las bacterias que contienen PR podría afectar la forma en que se procesa el DOC en el océano, la parte experimental de este trabajo se ha centrado en las respuestas de dos bacterias con sus genomas completamente secuenciados, *Polaribacter* sp. MED152 y *Dokdonia* sp. MED134, en diferentes condiciones de crecimiento. Por último, se hizo un intento para determinar la degradación de polímeros en un entorno natural: el Océano Ártico, de invierno a primavera, cuando los Bacteroidetes, entre otros grupos, se estaban convirtiendo en cada vez más activos en el uso de la materia orgánica.

RESUM

El fílum dels Bacteroidetes és un dels grups més abundants de bacteris a l'oceà. Representen una fracció significativa del bacterioplàncton marí, especialment en les zones costaneres, encara que estan globalment distribuïts. Els Bacteroidetes són molt eficients en la degradació de la matèria orgànica particulada, i prefereixen substàncies polimèriques en comptes de substàncies dissoltes. Contràriament a la majoria de grups de bacteris marins, molts representants tenen motilitat per reptació i tendeixen a ser més abundants adherits a partícules que en vida lliure a la columna d'aigua. Per tant, probablement presentin una estratègia diferent a la dels proteobacteris o cianobacteris i, en conseqüència, ocupin nínxols diferents. En conjunt, els Bacteroidetes juguen un paper únic i important en el cicle del carboni marí. En aquesta tesi s'ha utilitzat una varietat d'enfocaments per comprendre l'ecologia d'alguns membres d'aquest grup de bacteris.

En primer lloc, es va dur a terme l'anàlisi complert del genoma d'un representant dels Bacteroidetes (*Polaribacter* sp. MED152), del qual es van poder deduir estratègies de supervivència diferents a les d'altres grups. Posteriorment es va realitzar un anàlisi comparatiu *in silico* (fent servir eines bioinformàtiques) de quatre Bacteroidetes marins, els genomes han estat completament seqüenciats. Vam trobar que aquests bacteris tenien un gran nombre de gens implicats en la degradació de polímers i adhesió a superfícies. Un promig d'1,5% de la totalitat d'aquests genomes correspondria a gens adquirits per mitjà de transferència horitzontal de gens, proporcionava noves adaptacions i capacitats metabòliques. Això va portar a l'anàlisi d'illes genòmiques en bacteris marins, tractant de discernir possibles patrons característics als Bacteroidetes. Alguns dels bacteris analitzats tenien el gen proteorodopsina de la (PR), que els hi permet obtenir energia extra en presència de la llum. Atès que l'ús de la llum pels bacteris que contenen PR podria afectar la forma en què el DOC és processat a l'oceà, la part experimental d'aquest treball s'ha centrat en les respostes de dos bacteria, els genomes dels quals estaven completament seqüenciats, *Polaribacter* sp. MED152 i *Dokdonia* sp. MED134, en diferents condicions de creixement. Finalment, es va fer un intent per determinar la degradació de polímers en un entorn natural: l'Oceà Àrtic, d'hivern a primavera, quan els Bacteroidetes, entre altres grups, s'estaven convertint cada vegada més actius en l'ús de la matèria orgànica.

ABSTRACT

The Bacteroidetes phylum is one of the most abundant groups of bacteria in the oceans. Its members account for a significant fraction of marine bacterioplankton especially in coastal areas, although they are globally distributed. Bacteroidetes are successful in the degradation of particulate organic matter; they are claimed to prefer polymeric substances rather than small molecules. Contrary to most other groups of marine of bacteria, many representatives possess gliding motility and they tend to be more abundant attached to particles than free-living in the water column. Thus, they likely present a different strategy to that of Proteobacteria or Cyanobacteria and, therefore, they will fill different niches. Altogether, the Bacteroidetes are thought to have a unique and important role in the marine carbon cycle. The present dissertation uses a variety of approaches to understand the ecology of some members of this group of bacteria.

First, a complete genome analysis of a representative of the Bacteroidetes (*Polaribacter* sp. MED152) was conducted. This suggested a strategy for survival different from those of other groups. Then, a comparative *in silico* analysis (using bioinformatics tools) of four marine Bacteroidetes whose genomes had been completely sequenced was carried out. We found that these bacteria had a large number of genes involved in polymer degradation and adhesion to surfaces. From the analyzed representatives, an average of 1.5% of the whole genome corresponded to genes acquired by means of horizontal gene transfer, providing new adaptations and metabolic capabilities. This finding lead to the analysis of genomic islands in marine bacteria, trying to discern possible patterns characteristic of Bacteroidetes. Some of the bacteria analyzed had the proteorhodopsin gene (PR), allowing them to obtain extra energy in the presence of light. Since the use of light by PR-containing bacteria could affect the way DOC is processed in the ocean, the experimental part of the work focused on the responses of two whole genome-sequenced, PR-containing bacteria, *Polaribacter* sp. MED152 and *Dokdonia* sp. MED134, to different growth conditions. Finally, an attempt was made to determine the degradation of polymers in a natural environment: the Arctic Ocean from winter to spring, when the Bacteroidetes, among other groups, were becoming progressively more active in the use of organic matter.

LIST OF PUBLICATIONS

This thesis is based on the following papers:

1. González, J M., **B. Fernández-Gómez***, A. Fernàndez-Guerra*, L. Gómez-Consarnau*, O. Sánchez, M. Coll-Lladó, J. del Campo, L. Escudero, R. Rodríguez-Martínez, L. Alonso-Sáez, M. Latasa, I. Paulsen, O. Nedashkovskaya, I. Lekunberri, J. Pinhassi, and C. Pedrós-Alió. 2008. Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (Flavobacteria). *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*. 25: 8724-8729. (*Equal contribution).

   I contributed to this paper by manually annotating around 17% of the genes present in the genome of MED152 and subsequently I reconstructed several metabolic pathways: TCA cycle, pentose phosphate pathway, pyruvate metabolism, glycolisis, glyoxylate metabolism, Entner-Doudoroff pathway, and all amino acid biosynthesis pathways.

2. **Fernández-Gómez B**., M. Richter , M. Schüler, J. Pinhassi, S.G. Acinas, J. M. González, and C. Pedrós-Alió. Comparative genomics of marine Bacteroidetes with and without proteorhodopsin. Submitted to *BMC Genomics*.

3. **Fernández-Gómez B**.*, A. Fernàndez-Guerra*, E. O. Casamayor, J.M. González, C. Pedrós-Alió, and S.G. Acinas. Patterns and architecture of genomic islands in marine bacteria. (*Equal contribution). Submitted to *The ISME Journal*.

4. **Fernández-Gómez B**., J. M. González, and C. Pedrós-Alió. Culture medium influences the response to light of PR-containing marine Flavobacteria. Manuscript.

5. **Fernández-Gómez B**., M. M. Sala, and C. Pedrós-Alió. Seasonal changes in substrate utilization patterns by bacterioplankton in the Amundsen Gulf (Western Arctic). Submitted to *Journal of Marine Systems*.

THESIS PREVIEW /PRESENTACIÓN DE LA TESIS

La presente tesis, titulada *Ecología de los Bacteroidetes Marinos: Una aproximación genómica,* ha sido financiada por el Consejo Superior de Investigaciones Científicas (CSIC) a través de una beca predoctoral a Beatriz Fernández Gómez, en el programa I3P/Ventorrillo y asignada al proyecto: "Aproximación Genómica a la Ecología del Plancton Microbiano (GEμMAR)" en el Instituto de Ciencias del Mar bajo la supervisión de los Dres. Carlos Pedrós-Alió y José M. González (Ref. CTM2004-02586/MAR).

Otros proyectos que han contribuido a la finalización de esta tesis han sido: "Genómica y ecología de los microorganismos marinos (GEMMA)" (CTM2007-63753-C02-01/MAR, financiado por el Ministerio de Ciencia e Innovación, I.P. Dr. Carlos Pedrós-Alió) y "BOREAL. Participación en la campaña canadiense Circumpolar Flaw Lead Polynya Study" (CLG2007-28872-E/ANT, I.P. Dr. Carlos Pedrós-Alió).

This thesis, *titled Ecology of Marine Bacteroidetes: A genomics approach*, has been funded by The Spanish National Research Council (CSIC) through a PhD fellowship to Beatriz Fernández Gómez, under the program I3P/Ventorrillo and was ascribed to the project: "Aproximación Genómica a la Ecología del Plancton Microbiano (GEμMAR)" in the Institute of Marine Sciences under the supervision of Drs. Carlos Pedrós-Alió and José M. González (Ref. CTM2004-02586/MAR).

Other projects that contributed to the achievement of this thesis have been: "Genómica y ecología de los microorganismos marinos (GEMMA)" (CTM2007-63753-C02-01/MAR, funded by the Spanish Ministry of Science and Innovation, P.I. Dr. Carlos Pedrós-Alió) and "BOREAL. Participation in the Canadian Cruise Circumpolar Flaw Lead Polynya Study" (CLG2007-28872-E/ANT, P.I. Dr. Carlos Pedrós-Alió).

"Probably the greatest enticement for those who today are devoting their lives to the study of the sea is the lure of the unknown, the challenge of the undiscovered, the thrill of discovery on what is truly the last frontier on Earth (…) Only the ocean remains as the last great unexplored portion of our globe; so it is to the sea that man must turn to meet the last great challenge of exploration this side of outer space (…) But the influence of the sea on man's daily life and on his future well-being is so great and still so poorly comprehended that the sea must be explored, studied, and understood."

H. B. Stewart , *In Deep Challenge* (1966).

CONTENTS

RESUMEN/RESUM/ABSTRACT **15**

LIST OF PUBLICATIONS **19**

PRESENTACIÓN DE LA TESIS/THESIS PREVIEW **21**

INTRODUCTION **29**

OBJECTIVES OF THE THESIS **43**
METHODOLOGY **44**

Chapter 1. Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (Flavobacteria). **49**

Chapter 2. Comparative genomics of marine Bacteroidetes with and without proteorhodopsin. **91**

Chapter 3. Patterns and architecture of genomic islands in marine Bacteria. **149**

Chapter 4. Culture medium influences the response to light of PR-containing marine Flavobacteria. **195**

Chapter 5. Seasonal changes in substrate utilization patterns by bacterioplankton in the Amundsen Gulf (Western Arctic). **217**

SYNTHESIS OF RESULTS AND GENERAL DISCUSSION **241**

RESUMEN EN ESPAÑOL/SPANISH SUMMARY **261**

REFERENCES (Introduction, Discussion and Spanish Summary) AND ANNEXES **315**

"Nothing in life is to be feared,
it is only to be understood.
Now is the time to understand more,
so that we may fear less"
Marie Curie

General Introducction

ECOLOGICAL GENOMICS OF MARINE BACTERIA

If Life had a yearbook, bacteria would win all of the awards, especially "most likely to succeed". (...) By contrast, humans have occupied a narrow range of environmental conditions—and for only about 0.003 percent of the Earth's existence. If we even made it into the yearbook, the caption would have read "photo not available". **Lynn Margulis, "The Germs of Life".**

THE MARINE ENVIRONMENT

The oceans are the largest body of water on Earth. With 1.37×10^9 km^3 and approximately 3.61×10^{14} m^2, the oceans cover about 71% of the Earth's surface (Davis Jr., 1972). More than half of this area is over 3,000 meters deep, the deepest point being the Mariana Trench, close to 11,000 meters deep. Salinity of seawater is in the range of 33 to 37 with an average around 35. The pH at the surface averages 8.1 and it is becoming more acidic since the industrial revolution began (Orr *et al*., 2005). The average temperature of the ocean surface waters is about 17°C, changing mainly with latitude. The polar seas can be as cold as -2°C while the Persian Gulf can be as warm as 36°C.

Oceans significantly influence the Earth's climate (Stewart, 2008) by redistributing heat from the tropics to the polar regions (wind-driven and thermohaline circulation), exchanging heat and water vapor with the atmosphere (influencing rainfall patterns and atmospheric circulation), and acting both as sources and sinks of CO_2. Disruptions in the thermohaline circulation are thought to have considerable impacts on the Earth's climate (Bryden *et al*., 2005).

The marine environment contains a vast amount of carbon. One part of it is present in the form of organic carbon, either as dissolved organic carbon (DOC) or as particulate organic carbon (POC). Carbon fuels life in the oceans as it is fixed through photosynthesis by phytoplankton, providing the basis of the marine food webs. Despite their relatively low concentrations, nutrients such as nitrogen (N) and phosphorus (P) are particularly important to the biology of the oceans, and they are the limiting nutrients in most productive regions. In some cases, iron (Fe) or silica (Si) may also act as limiting nutrients (particularly in open ocean regions). The ratio of the concentrations of carbon to nitrogen to phosphorous in phytoplankton is 106:16:1, which is the classic "Redfield ratio" (Redfield, 1934).

BACTERIOPLANKTON AND ITS ROLE IN THE OCEAN

Life within the oceans evolved at least 3 billion years earlier than on land. Microbes were the only kinds of life on Earth for around 80% of the Planet's history and all multicellular life depends upon microbial processes (Amaral-Zettler *et al*., 2010). Bacterioplankton includes the bacterial and archaeal members of the plankton (Thurman, 1997), comprising the smallest and most abundant organism in the ocean (Whitman *et al*., 1998). The ecological function of marine bacteria and archaea is crucial for life on Earth by controlling the ecology of the marine environment as producers and consumers of dissolved organic matter (DOM) (Kirchman, 2008). They are the oldest life forms, the primary catalysts of energy transformations, and fundamental to the biogeochemical cycles that shape our Planet (Falkowski *et al*.,

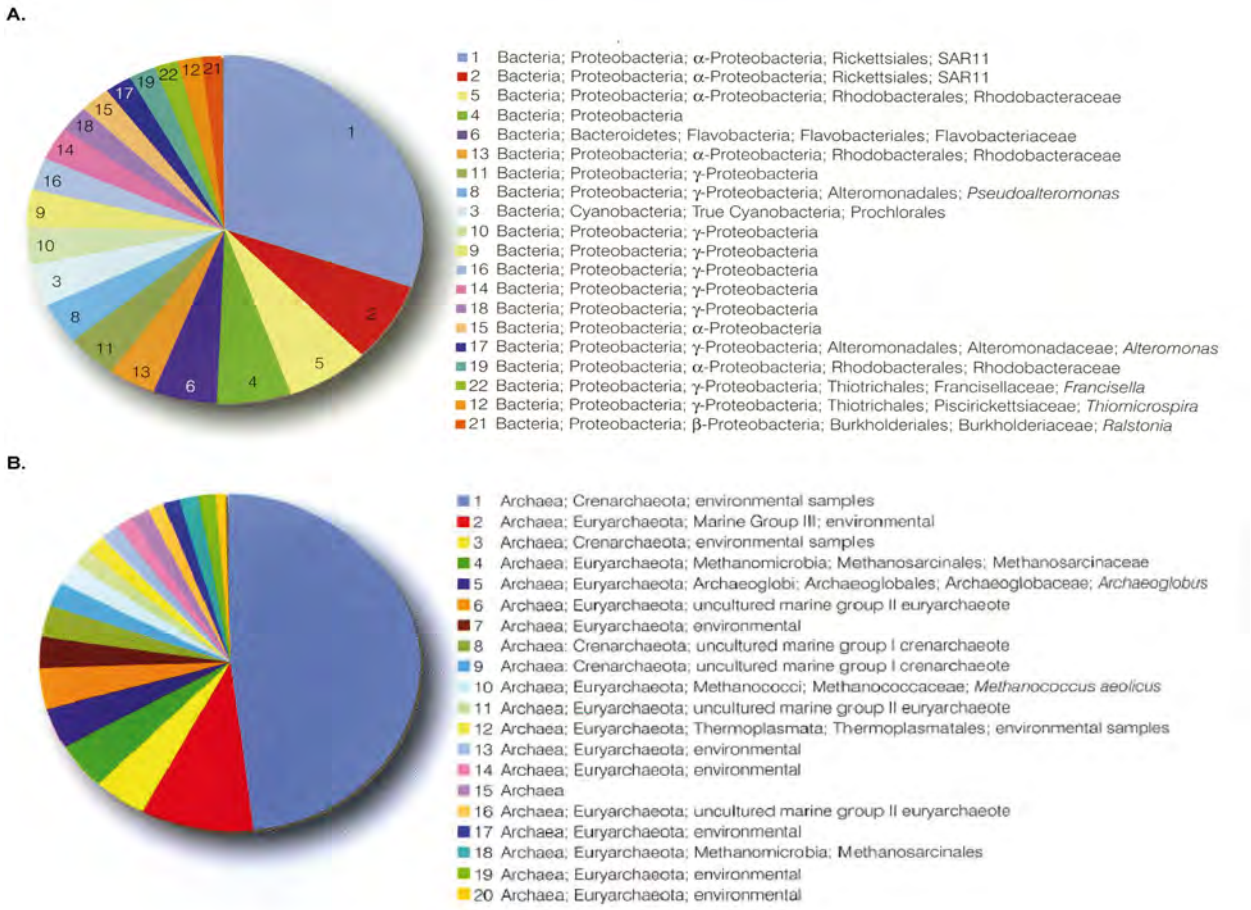


Figure 1. (A) Taxonomic breakdown of the top 20 most abundant bacterial sequences found across 583 bacterial datasets. (B) Taxonomic breakdown of the top 20 most abundant archaeal sequences found in 120 bacterial datasets. The rankings for both pies are based on the sum of the relative abundances of individual sequences from each sample. Taxonomies are base on the Global Alignment for Sequencing Taxonomy (GAST) procedure (Huse *et al.*, 1997). Figure from A Global Census of Marine Microbes (Amaral-Zettler *et al*., 2010).

2008). They are remarkably abundant, around 10^5 cells per cm^3 (Hobbie *et al*., 1977; Porter and Feig, 1980; Zimmerman R. and Meyer-Reil, 1974) and when extrapolating to the whole ocean, considering a volume of $1.4x10^9$ km^3, they amount to approximately 10^{29} prokaryotic cells (Whitman *et al*., 1998). To put this number in perspective, one can consider that there are only 10^{11} neurons in a human brain or stars in the galaxy.

The ocean harbors numerous and diverse groups of bacteria and archaea (Fig. 1). Several of them are autotrophs, including the phytoplankton, and derive energy from photosynthesis. Primary producers in the open ocean are responsible for about half of the Planet's total primary production (Falkowski *et al*., 1998; Field *et al*., 1998). They comprise the Cyanobacteria, that include the two most abundant genera in the ocean: *Prochlorococcus* (Waterbury *et al*., 1979) and *Synechococcus* (Chisholm *et al*., 1988). Nevertheless, the largest part of marine bacteria consists of heterotrophs that acquire energy by consuming DOM and POM, mostly derived from phytoplankton (Fogg, 1958; Sharp, 1977) but also from grazers (Nagata, 2000). Viruses also contribute to the OM pool by inducing cell lysis (Wilhelm and Suttle, 1999). Heterotrophic bacteria oxidize approximately one-half of the labile carbon fixed by photosynthesis (Azam *et al*., 1983b; Fenchel, 1988; Robinson, 2008), although this proportion can be significantly higher in oligotrophic waters (del Giorgio *et al.,* 1997). As a consequence, carbon dioxide is released to the ocean and a portion of the organic matter is remineralized into essential inorganic nutrients. This biomass can be reintroduced in the

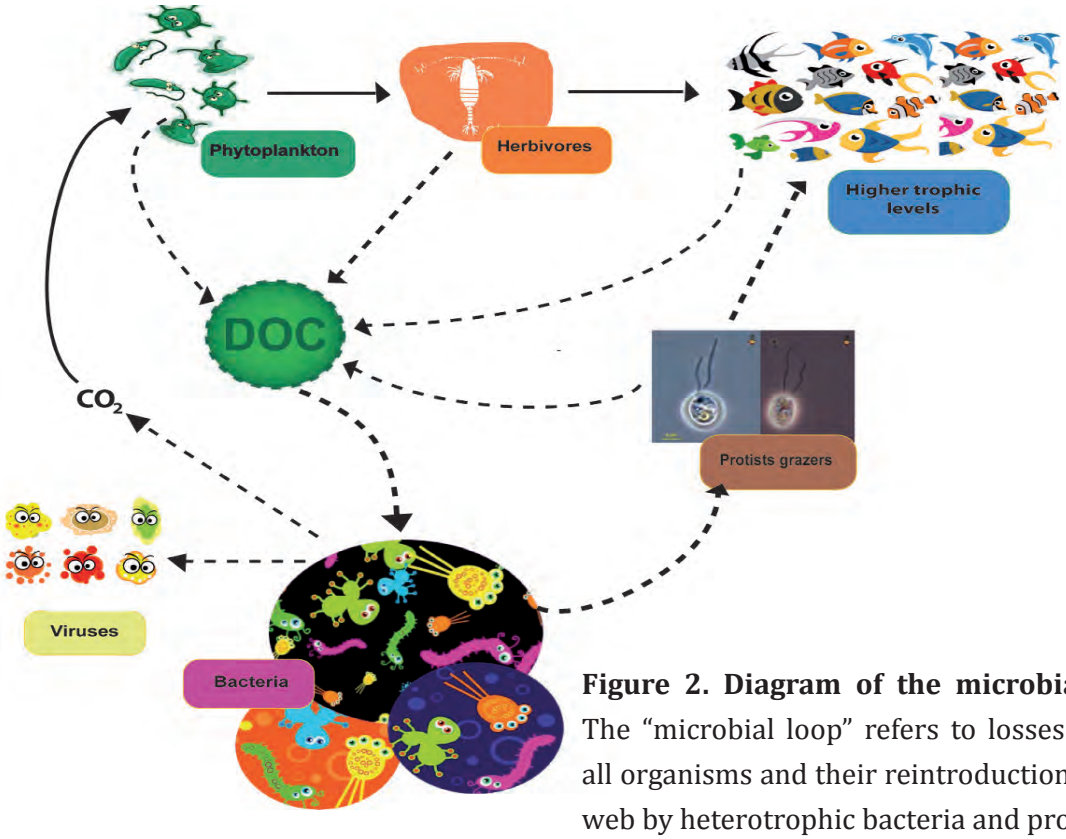


Figure 2. Diagram of the microbial food web. The "microbial loop" refers to losses of DOC from all organisms and their reintroduction into the food web by heterotrophic bacteria and protists.

food web and represents an additional source of organic substrates for higher trophic levels (protists), while limiting the loss of organic matter via sedimentation and deep burial. Hence, bacteria play a crucial role in pelagic food webs by controlling much of the carbon fluxes in the open ocean (Azam, 1998). This particular food chain is known as the "*microbial loop*" (Azam *et al*., 1983a; Ducklow, 1983; Pomeroy, 1974) (Fig. 2).

At present, with increasing levels of atmospheric CO_2, understanding the function of microbes in marine food webs is an essential issue since the *microbial loop* determines, in part, the response of oceanic ecosystems and the carbon cycle to climate change (Falkowski *et al*., 2000; Kirchman *et al*., 2009).

Bacterioplankton also play significant roles in other biogeochemical cycles such nitrogen, phosphorous, or sulfur and anaerobic processes such as fermentation (Kirchman, 2000).

MARINE HETROTROPHIC BACTERIA: The phylum Bacteroidetes

Different groups of heterotrophic bacteria are present in the ocean (Fig. 1). The most abundant groups belong to the phylum Proteobacteria, particularly the classes Alphaproteobacteria and Gammaproteobacteria (Giovannoni and Rappé, 2000), which represent 30-40% of the prokaryotic cells; and the phylum Bacteroidetes, representing between 10-30% (Cottrell and Kirchman, 2000a; Glöckner *et al*., 1999). The phylum Bacteroidetes, formerly know as Cytophaga-Flavobacterium-Bacteroides (CFB) group, is diverse and is composed mainly by three large classes of bacteria widely distributed in the environment, including soils, sediments, hydrothermal vents, fresh- and sea water, and the gut and skin of animals: Bacteroidia, Flavobacteria, and Sphingobacteria (Fig. 3 and Fig. 4).

Class Bacteroidia is the best studied in general microbiology because of its implications for human health (Finegold *et al*., 1983). It includes the genus *Bacteroides* (the most abundant in the intestine of warm-blooded animals including humans) and *Porphyromonas*, a group of organisms inhabiting the human oral cavity. However, the Flavobacteria and Sphingobacteria classes, whose representatives are mostly from marine and soil habitats, have only received attention in the last decades.

Flavobacteria, the most abundant class of the Bacteroidetes phylum in the sea (Alonso *et al*., 2007), are assumed to be specialized in the degradation of polymers and particulate organic matter (POM) (Pinhassi *et al*., 2004; Riemann *et al*., 2000). Many representatives have gliding motility and are important inhabitants of marine aggregates (DeLong *et al*., 1993; Kirchman, 2002), having the capability to use polymeric substances as carbon and

energy sources.

During algal blooms Flavobacteria members are overrepresented with respect to other bacteria as a response to the organic carbon present, denoting their predilection for large molecules, which corresponds to their assumed role as major mineralizers of organic matter (Cottrell and Kirchman, 2000b).

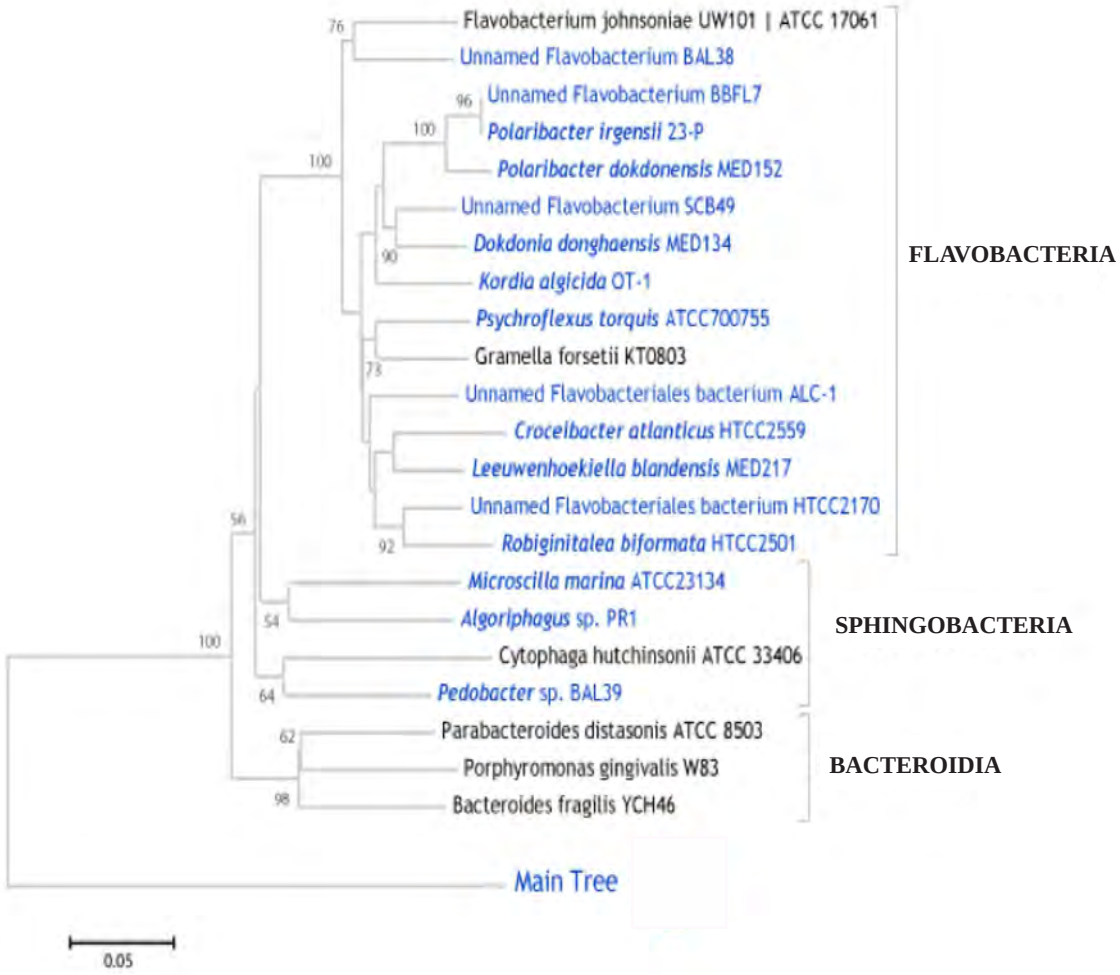


Figure 3. Phylogenetic tree of the main classes within the phylum Bacteroidetes (G. & B. Moore Foundation webpage).

This feature contrasts with that of the other major group of marine bacteria: the Proteobacteria. Both Alpha- and Gammaproteobacteria seem to prefer monomers rather than polymers and carry out a free-living existence in the water column, a fact that suggests complementary roles for the complete degradation of organic matter (Cottrell and Kirchman, 2000b).

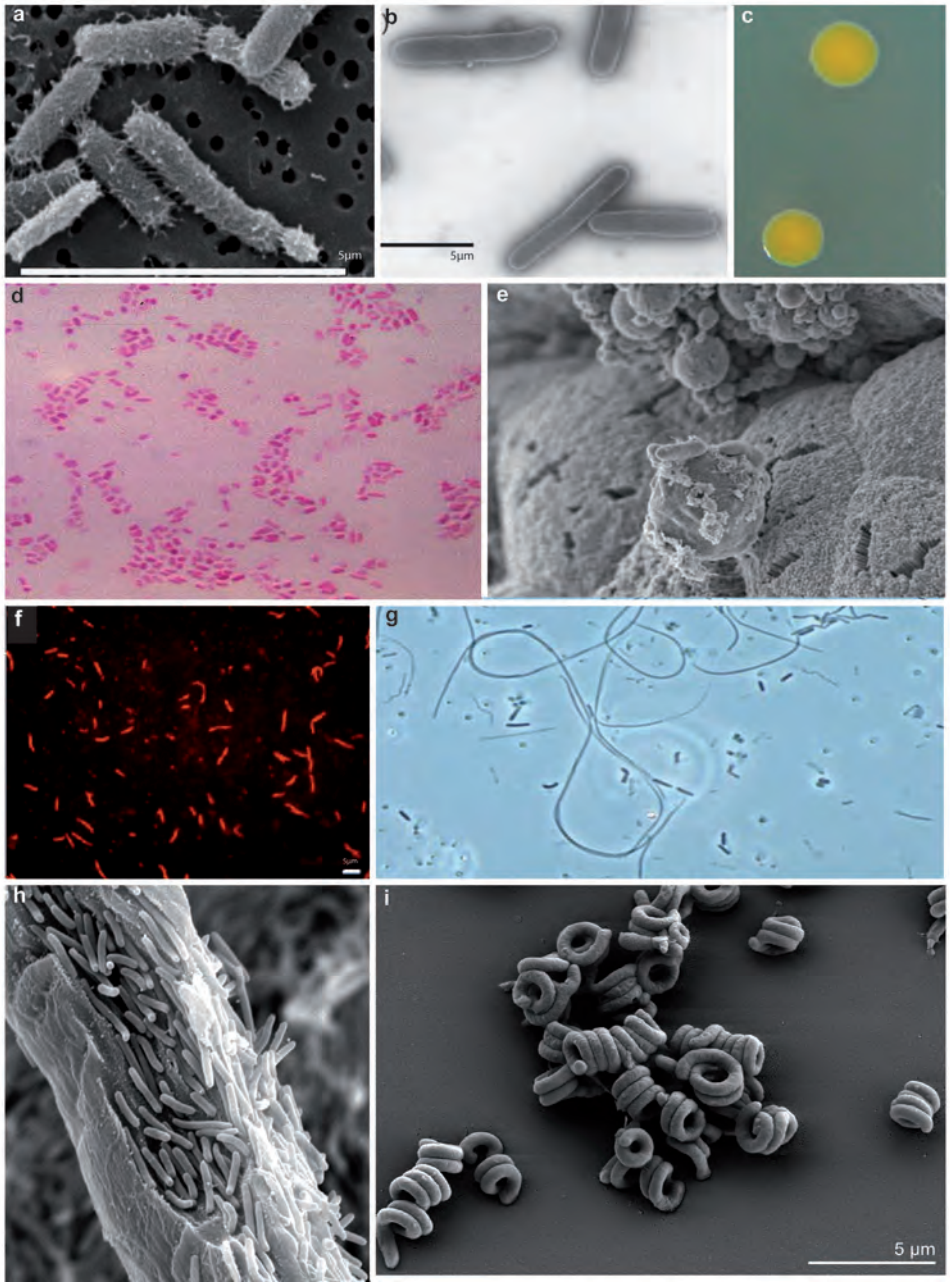


Figure 4. Diversity of marine Bacteroidetes. a-c: Flavobacteria; d-e: Bacteroidia; f-i: Sphingobacteria. a. Scanning electron microscopy image of cells of *L. blandensis* MED217 in the exponential growth phase. **b.** Electron micrographs of negatively stained cells of *Robiginitalea biformata* HTCC2501 in exponential phase. **c**. Colonies of *Flavobacterium psychrophilum*. **d.** Gram strain of *Bacteroides fragilis*, an obligate gut microbe. **e.** *Bacteroides thetaiotaomicron* in association with food particle in the mouse gut. **f.** Identification by FISH of *Salinibacter ruber.* **g.** *Flexibacter sp.* **h**. Scanning electron micrograph of the cellulolytic gliding bacterium *Cytophaga hutchinsonii*, showing cells gliding on and digesting a cellulose fiber. **i.** Scanning electron micrograph of *Spirosoma linguale*.

Despite their abundance, Flavobacteria have received much less attention than Proteobacteria and Cyanobacteria.

Thus, the main objective of this dissertation was to study the ecology and physiology of some representatives of the Flavobacteria.

MAIN CARBON SOURCES FOR BACTERIOPLANKTON

Although it is known that seawater contains organic matter in a size continuum of discrete units (Sharp, 1973) and these units are interconnected (Verdugo *et al.*, 2004), organic matter in the ocean has traditionally been divided into two groups (dissolved and particulate) pragmatically based on filtration through GF/F filters (Fig. 5).

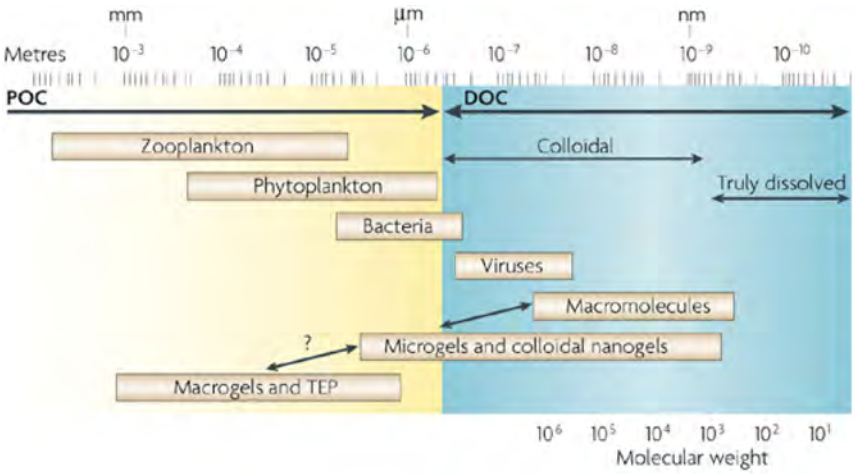


Figure 5. The size range of organic matter and microbial interactions in the ocean (Azam and Malfatti 2007).

The main form of organic carbon (>97%) in seawater is represented by DOM (Benner, 2002) which is mainly produced by algal exudation (Nagata, 2000) and cell lysis (viral infections) (Weinbauer and Peduzzi, 1995). However, bacteria also produce cell-surface mucus that can be released to the medium. A large fraction of this DOM consists of refractory molecules that bacteria are not able to assimilate, and it may remain in the ocean for thousands of years (Bauer *et al.*, 1992; Bauer, 2002; Williams and Druffel, 1987). But there is also a small fraction (<10%) of available (labile) molecules (mono- and polysaccharides, proteins, peptides and aminoacids) (Benner, 2002) that bacteria can easily utilize. DOM is mainly released in the form of high-molecular-weight DOM (HMW-DOM, >1kDa), which is preferred by bacteria for its high bioreactivity compared with refractory low-molecular-weight-DOM (Amon and Benner, 1994). However, bacteria can only incorporate small molecules into the cell through transmembrane channels. In order to assimilate large molecules they first need

to break them by means of specific extracellular hydrolytic enzymes such as glucosidases, peptidases, lipases or chitinases. (Arnosti *et al*., 2005; Hollibaugh and Azam, 1983; Martínez *et al*., 1996).

Organic matter that is retained on glass-fiber filters of approximately 0.7μm of pore size constitutes POM. POM is generally more biodegradable than DOM (Amon and Benner, 1996; Hedges, 1992). POM includes both living (plankton) and non living (detritus) components (Nagata, 2008). The majority of POM consists of small particles in the form of gels, floating in seawater, that originate from the aggregation of marine colloids (which represent 10-40% of DOM) (Chin *et al*., 1998; Kepkay, 1994). These macromolecules tend to aggregate forming macro-gels that eventually may reach the size of sinking particles (Engel *et al*., 2004). When the aggregates have macroscopic size (>500μm) they are known as "*marine snow*" (Suzuki N. and Kato K., 1953). *Marine snow* refers to the aggregation of large particles and the agglomeration with other materials such as phytoplankton or detritus. It has been shown that these aggregates harbor bacterial communities that differ from those in the water column (Alldredge *et al*., 1986). Bacteria attached to *marine snow* posses high hydrolytic enzyme activity (Smith *et al*., 1992), leaving an extended plume of DOM behind, that other bacteria can track.

Marine Flavobacteria are especially proficient at degrading HMW-DOM (Cottrell and Kirchman, 2000b; Cottrell *et al*., 2005; Kirchman, 2002) and they are found often associated to *marine snow* rather than free-living (Alldredge *et al*., 1986; DeLong *et al*., 1993; Johannes *et al*., 1998). Due to the implications of colonizing *marine snow* for the carbon cycle, marine Bacteroidetes have a substantial role in the structuring of the marine ecosystem.

In the present work I analyzed the genomes of several Flavobacteria, looking for potential degradative enzymes (glycosidases and peptidases), as well as other genes that might confirm their predilection for macromolecules and their association with aggregates. The purpose of this exercise was to test whether the role postulated for these bacteria in the already mentioned studies was coherent with their genetic endowment.

PHOTOHETEROTROPHY IN THE OCEAN: Proteorhodopsins

Among the heterotrophic bacteria, several groups include members with mechanisms to obtain energy from sunlight, a metabolic process known as photoheterotropy. Two major light-harvesting systems allowing photoheterotrophy are known in the ocean: bacterioclorophyll *a* in aerobic anoxygenic photosysnthetic bacteria (AAPs) and proteorhodopsins (PR), in representatives of ß-proteobacteria (i.e *Methylophilales* sp.),

α-proteobacteria (i.e *Pelagibacter* sp.), γ- proteobacteria (i.e *Vibrio* sp.), Flavobacteria (i.e *Dokdonia* sp.), and Planctomycetes (McCarren and DeLong, 2007).

Proteorhodopsins are bacterial rhodopsins, a family of transmembrane retinal-binding proteins that function as light-driven proton pumps (Béjà *et al*., 2000). They are homologs to the previously discovered bacteriorhodoposins in archaea from saltern cystallizer ponds (Oesterhelt and Stoeckenius, 1973). Both kinds of rhodopsins work as light-driven proton pumps that generate a chemiosmotic membrane potential from light energy, eventually resulting in energy storage as adenosine triphosphate (ATP) (Fig. 6).

Their discovery was a genuine revolution, demonstrating that more metabolisms existed than the previously known phototrophic metabolisms based on chlorophyll-like molecules in the sea. They were discovered after the sequence analysis of a cloned genome fragment from an uncultivated marine SAR86 clade member (Béjà *et al*., 2000).

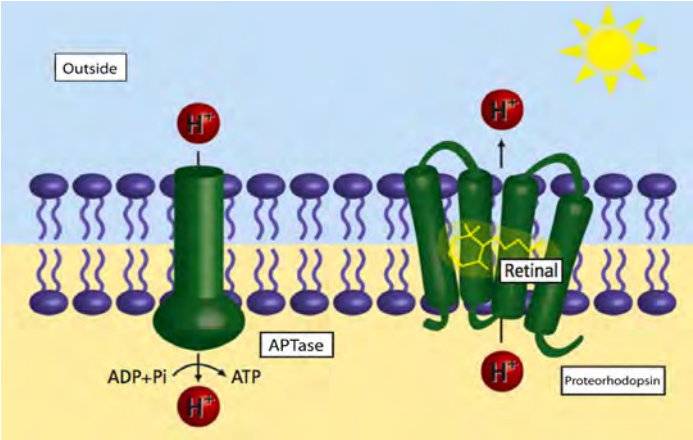


Figure 6. Outline of the predicted system of PR-dependent ATP synthesis. José M. González (Investigación y Ciencia).

Successive studies detected a wide range of PR gene variants and distributed among numerous marine bacteria (Campbell *et al*., 2008; de la Torre *et al*., 2003) and archaea (Frigaard *et al*., 2006). The marine microbial genomic sequencing initiative of the Gordon and Betty Moore Foundation (www.moore.org) allowed to sequence ~170 genomes of marine bacteria. The analysis of these genomes revealed the presence of PR in several bacteria such as Gammaproteobacteria (i.e. *Vibrio campellii* AND4, *Vibrio angustum* S14), Alphaprotebacteria (i.e. *Photobacterium* sp. SKA34, *Pelagibacter ubique* HTCC1062), Betaproteobacteria (i.e Methylophilales bacterium HTCC2181) and Bacteroidetes (i.e *Dokdonia* sp. MED134, *Polaribacter* sp. MED152, *Polaribacter irgensii* 23-P, Flavobacteria bacterium BAL38).

Although it was assumed that possessing the PR gene provided an adaptive advantage, there was no clear evidence regarding the metabolic role of PR, mainly due to the fact

that only a few bacteria with PR had been successfully cultured (Fuhrman *et al*., 2008; Riedel *et al*., 2010).

Only *Dokdonia* sp. MED134, isolated from the Blanes Bay Microbial Observatory (BBMO), has been shown to have higher yields in the light than in the dark when limited by organic matter (Gómez-Consarnau *et al*., 2007; Kimura *et al*., 2011). This means that MED134 could get energy from light during oligotrophic conditions and survive until it were able to colonize new particles to attach to. It remains to be proven whether this occurs in natural communities from oligotrophic waters.

In the present work I investigated whether possessing the PR could actually benefit bacteria under conditions of low concentration of organic matter.

MICROBIAL GENOMICS versus CULTIVATION

Since the so-called "genomics revolution" started about 15 years ago, microbial genomics and metagenomics have become powerful tools to explore the microbial world (Xu, 2006). These tools have the potential to revolutionize our understanding of the role of microorganisms in the ocean (Pedrós-Alió, 2006).

The study of all genes encoded in the genetic material possessed by an organism became possible thanks to the technological developments in the field of molecular biology (Moran *et al*., 2004a). At the beginning, the capacity for sequencing allowed determination of the diversity and distribution of selected genes. For instance, the amplification and sequencing of the 16S rRNA or *nif*H genes provided information on the phylogenetic relationships among species (Woese *et al*., 1975) or the nitrogen-fixing metabolism of bacteria (Zehr and Capone, 1996) respectively.

These molecular studies differ from genomics in the fact that the latter allows to analyze more genes at a time and the possibility of finding novel genes without having to know them previously. It is estimated that less than 1% of all bacterial species have been cultured (Amann *et al*., 1990). Thus, the cultured microbes provide a window into the study of their uncultured relatives, becoming model organisms to test hypotheses about their physiology. Some examples of marine model microbes are: *Prochlorococcus marinus* (Rocap *et al*., 2003), *Pelagibacter ubique* (Giovannoni *et al*., 2005), *Silicibacter pomeroyi* (Moran *et al*., 2004b), or *Rhodopirellula baltica* (Glöckner *et al*., 2003). Genomics of cultured microorganisms is clarifying who is doing what and how they are doing it. However, genomics have to face the impossibility of having all microbes in pure cultures.

Metagenomics is distinct from, but complementary to, genomics of isolates. Also re-

ferred to as community or environmental genomics, metagenomics is the sequencing and analysis of DNA of all the microorganisms recovered from an environment, bypassing the need to isolate and culture individual microbial species (Chen and Pachter, 2005). The resultant wealth of genes deciphered from uncultured microorganisms has tremendous potential in the development of novel biocatalysts for industrial and medical applications. Two examples of metagenomics surveys are the Sargasso Sea study (Venter *et al*., 2004) where a total of 1.045×10^9 base pairs of sequence were generated and the Global Ocean Sampling (Rusch *et al*., 2007) that yielded an extensive dataset consisting of approximately 6.3×10^9 bp of sequences. Another example is the analysis of clone libraries with large fragments of DNA in each clone. It was with this approach that the PR genes were discovered (Béjà *et al*., 2000). Other studies revealed the presence of an ammonium monooxygenase gene (for ammonium oxidation, producing nitrite) in archaea, forcing a reevaluation of the preconception that oceanic nitrification was an exclusively bacterial function (Nicol and Schleper, 2006).

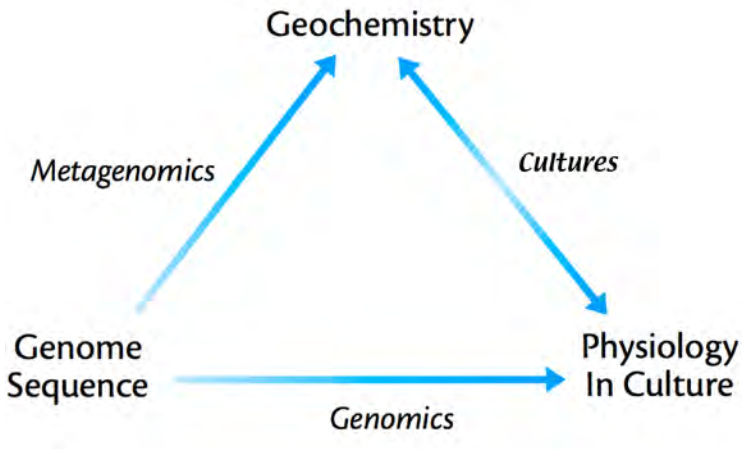


Figure 7. Relationships between genomics, metagenomics, and culturing (adapted from Giovannoni *et al*., 2007).

Despite the substantial accomplishments of metagenomics in predicting microbial metabolic activities from genome sequences, culturing remains an essential technique that allows the study of physiological characteristics of the microorganisms. The study of microorganisms in culture is the only way to test hypotheses generated by *in silico* analysis of genome sequences or ecological studies (Giovannoni *et al*., 2007). Figure 7 shows the relationships between the different disciplines mentioned in this paragraph: Genomics increases the amount of information retrieved from a genome, enhancing the number of targeted physiological studies that can be done in the laboratory. These physiological studies, in turn, increase the quality of predicting the geochemical activities of cells in nature. Metagenomics avoids the need for cultures but then, the hypotheses arising from the genomes are more difficult to test. The advantage is that metagenomics reveals what is there without the need to know it in advance.

Most of the present thesis was done using genomics of marine isolates as a way to investigate and compare the genomic potential of Bacteroidetes. Culturing was an attempt to corroborate the information retrieved from the genomes.

HORIZONTAL GENE TRANSFER GIVES RISE TO GENOMIC ISLANDS

One mechanism of bacterial evolution is the acquisition of new genes, that might be beneficial under certain environmental conditions, through horizontal gene transfer (HGT) (Dobrindt *et al*., 2004). This mechanism allows bacterial populations to quickly diversify and adapt to new ecological niches (Boucher *et al*., 2003). Many of the genes acquired by HGT form clusters known as genomic islands (GIs). GIs are DNA fragments flanked by elements (like transposases or integrases) that allow the mobilization of the genetic material. GIs comprise mobile DNA elements such as integrative and conjugative elements and/or prophages (Fig. 8). These islands are characterized by relatively large size (Hacker and Kaper, 2000), frequent association with tRNA-coding genes (Williams, 2002), and a different percentage in GC content compared with the rest of the genome (Navarre *et al*., 2006). When they were first discovered, these elements were associated with pathogenicity (Hacker *et al*., 1990), but currently it is accepted that GIs represent a much broader variety of functions than pathogenicity (Dobrindt *et al*., 2004). GIs harbor a large diversity of genes in which their structure and origin varies among the genomes. However, some general features and patterns can be recognized through comparisons of different bacterial genomes (Vernikos and Parkhill, 2008). A few studies describing GIs of environmental microbial genomes suggest that the acquisition of additional metabolic traits such as defense mechanisms, transport of certain substances or the possibility to degrade metals, are common traits that characterize the GIs.

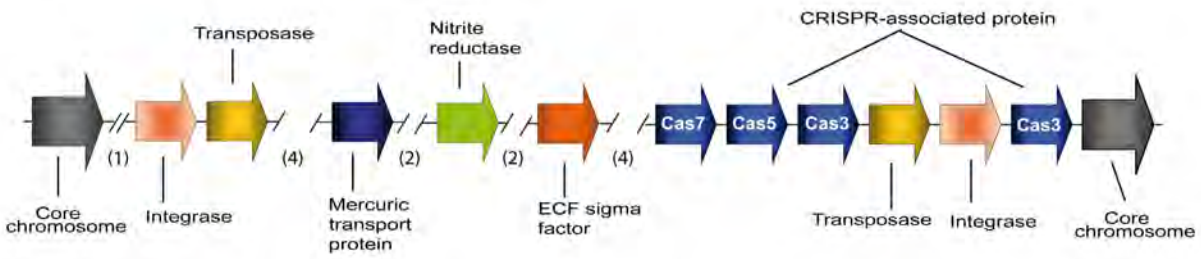


Figure 8. Schematic model of a genomic island of bacteria. Numbers in brackets indicate hypothetical proteins not considered for the figure.

In order to study these GIs, genomic island prediction software has been developed. Sequence based methods depend on the naturally occurring variation that exists between the genome sequence composition, for example, regions where the GC content changes, but they may have difficulty detecting GIs due to the normal variation in sequence composition that can occur within bacterial genomes. New approaches like comparative genomics based methods try to identify regions horizontally transferred using information from several related species (Langille *et al*., 2010).

In the present thesis, I determined the genomic islands of some representatives of the four major groups of bacteria in the ocean and analyzed them, searching for patterns and genes contained within these genomic islands that could provide better adaptations to the environment.

OBJECTIVES OF THE THESIS

The overall goal of this thesis was to characterize one important group of marine bacteria both genetically and physiologically, and to try to understand their life strategies in the ocean. The specific objectives of this thesis are the answers to the following questions:

- How is the genome of *Polaribacter* sp. MED152 ? What does the analysis of its genome tell us about the life strategy of this organism? **Chapter 1**

- Is there any pattern in the genome composition among marine Bacteroidetes? Could similar characteristics to those of *Polaribacter* be found in other representatives? **Chapter 2**

- Do the genomic islands present in some groups of bacteria provide advantages for ecological fitness? Are those different from other groups of bacteria? **Chapter 3**

- What is the actual function of the predicted genes (and more specifically the PR gene) in their environment? Can the traditional pure cultures techniques help to clarify this point? **Chapter 4**

- Is it possible to determine the metabolic capabilities of bacterioplankton in a natural environment? **Chapter 5**

METHODOLOGY

In the present work, bioinformatics tools have been used to reach most of the objectives cited above (**Chapters 1-3**) *in silico*. Culture experiments have also been carried out *in vivo* (**Chapter 4**) in order to test some of the hypotheses emerged from the former genomic analysis. Finally, a metabolic fingerprint technique (Biolog microplates) was used to characterize the microbial community *in situ* (**Chapter 5**). Although the particular methodologies are detailed in each chapter, a general overview of the different techniques is presented here.

The following diagram shows the different approaches used in the study of marine bacteria and their interconnections.

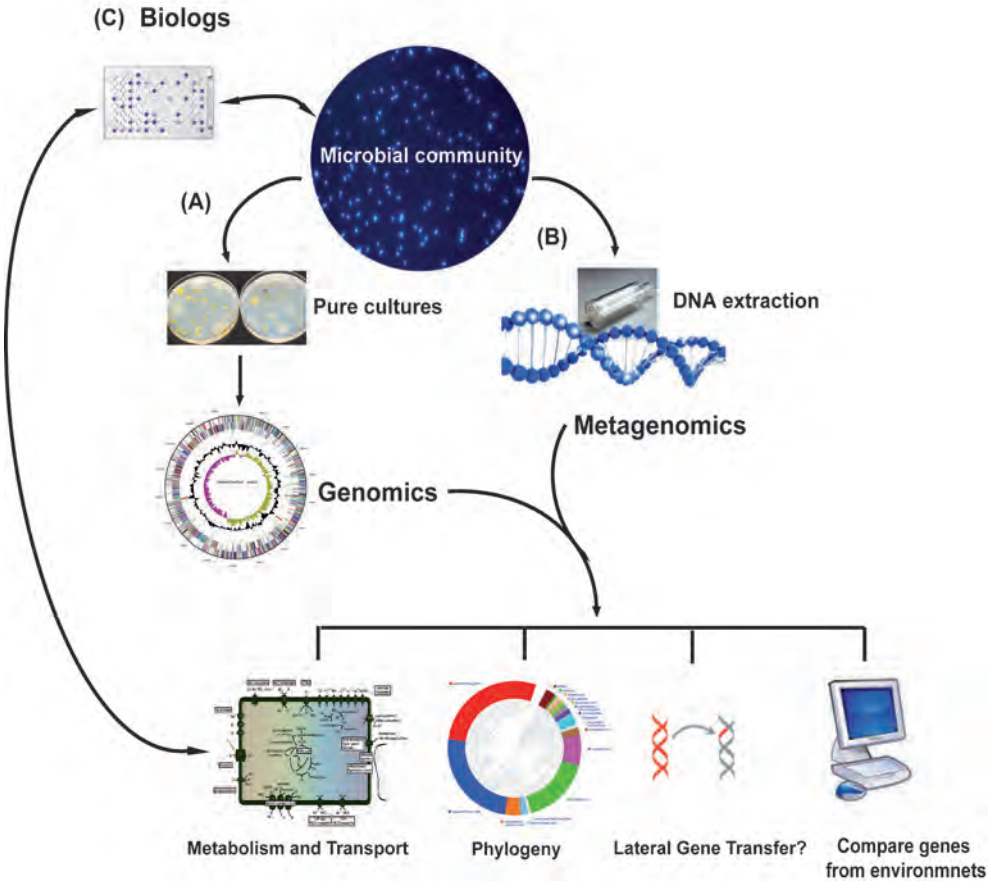


Fig. 9. Different approaches to study the microbial community used in this thesis. **(A)** An organism can be isolated in pure culture. Sequencing of its genome allows to study DNA, mRNA, and proteins to obtain information about: i) phylogeny, ii) metabolism and transport, iii) existence of lateral gene transfer, and iv) similarities with organisms from other environments. **(B)** Similar information can be obtained studying the entire microbial community through metagenomics, bypassing the need to culture the organisms (in the present thesis chose pathway **A**). **(C)** BIOLOG microplates allow studying some metabolic capabilities in samples taken directly from the community. Some of the traits found by means of genomics and metagenomics can also be tested using these plates.

In silico approaches: BIOINFORMATICS

Genome annotation and analysis

Genome annotation is the process of assigning biological information to gene sequences and their protein products (Stein, 2001). During the last years a large number of genomes have been sequenced (Overbeek *et al*., 2004) and manual annotation of the genomes is a very time consuming task. To solve this bottle neck, automatic annotation pipelines have been developed. Pipelines perform similarity searches, followed by automatic evaluation of the results and generation of functional annotation. Automatic annotation speeds up the process. However, a high quality genome annotation relies on the accuracy of the annotations, and manual curation is necessary. There are several portals that offer the tools and services for automatic annotation and curation. In this thesis **(Chapter 1)**, we used GenDB from the Center for Biotechnology (CeBiTec), at the University of Bielefeld (Germany) for this purpose.

Figure 10 shows the data flow in genome annotation. The process starts with the finished sequence of the genome and the automatic prediction of genes and their function. There is a feedback from gene identification to correction of sequencing errors. After the automatic prediction of genes, a similarity search done with the **Basic Local Alignment Search Tool** (BLAST) (Altschul *et al*., 1990) identifies the sequences with high similarity with other proteins in data bases (typically NCBI NR), and, therefore, their function can be predicted. A more detailed annotation can be carried out using other tools in combination with specialized databases such as KEGG (Kanehisa and Goto, 2000) or Pfam (Finn *et al*., 2008).

Genome comparison

Comparative genomics is the study of the relationships of genome structure and function among different organisms. While analysis of a single genome provides tremendous biological insights on any given organism, comparative analysis of multiple genomes provides significantly more information on the physiology and evolution of microbial species and expands the capacity to better assign putative function to predicted coding sequences. Comparative analysis offers a qualitatively new perspective on homologous relationships between genes (Koonin *et al*., 2005). By comparing the sequences of all genes between genomes and within each genome it is possible to reconstruct the evolutionary history of each gene. This, in turn, allows to better understand the specific adaptations of the genomes (Koonin *et al*., 2005).

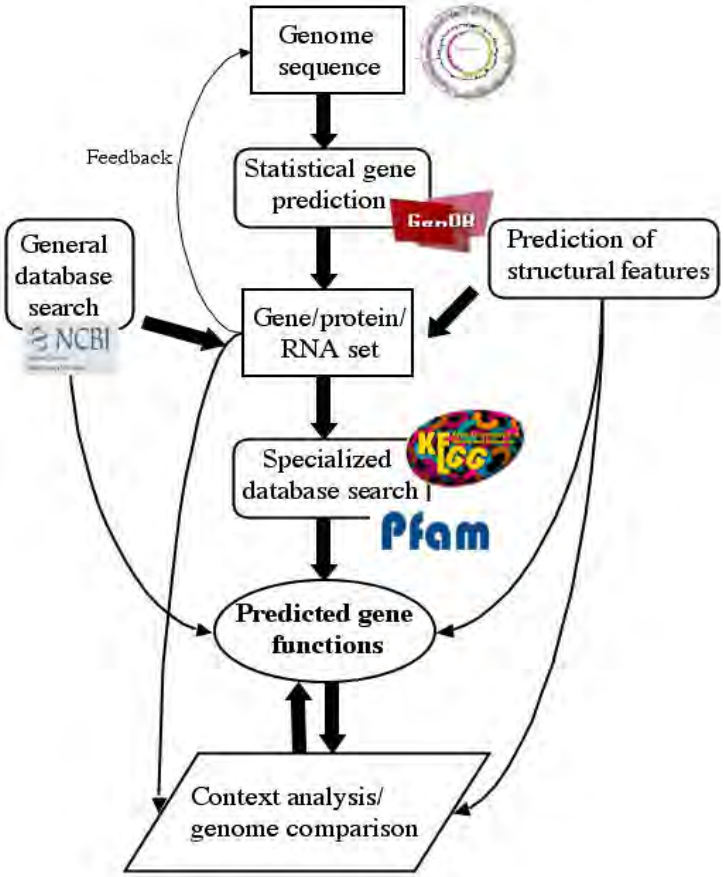


Figure 10. Generalized flow chart of genome annotation.

Comparative genomics involves the use of computer programs that can line up numerous genomes and look for regions of resemblance among them. Some of these sequence-similarity tools are accessible to the public over the Internet, the most widely used is the already mentioned BLAST. In **Chapter 2** we used the so-called **Reciprocal Best Match** (RBM) criterion to find orthologs (homologous genes with the same function related via speciation) and paralogs (homologous genes related via duplication and evolving new functions) among and within different marine Bacteroidetes representatives. This is nothing but BLAST performed in two ways, gene *a* against gene *b* and viceversa and if the similarity is good enough in both comparisons (according to the parameters defined) they are then considered to be homologous.

In order to explore and easily visualize the results, the software JCoast, www.jcoast.net (Richter *et al.*, 2008) was used, allowing to quickly find a gene of interest and compare it with its orthologs in other genomes.

Prediction of Genomic Islands

It is possible to detect genomic islands (GIs) within genomes thanks to the several programs that have been developed recently. In **Chapter 3** we chose IslandViewer, an integrated interface for computational identification and visualization of genomic islands (Langille and Brinkman, 2009; Langille *et al*., 2010). IslandViewer incorporates three different methods: SIGI-HMM (measures codon usage), IslandPath-DIMOB (measures dinucleotide bias), and IslandPick (automated comparative genomics based method). This web site contains pre-computed results for a large number of sequenced genomes and allows the user to submit genomes to be analyzed.

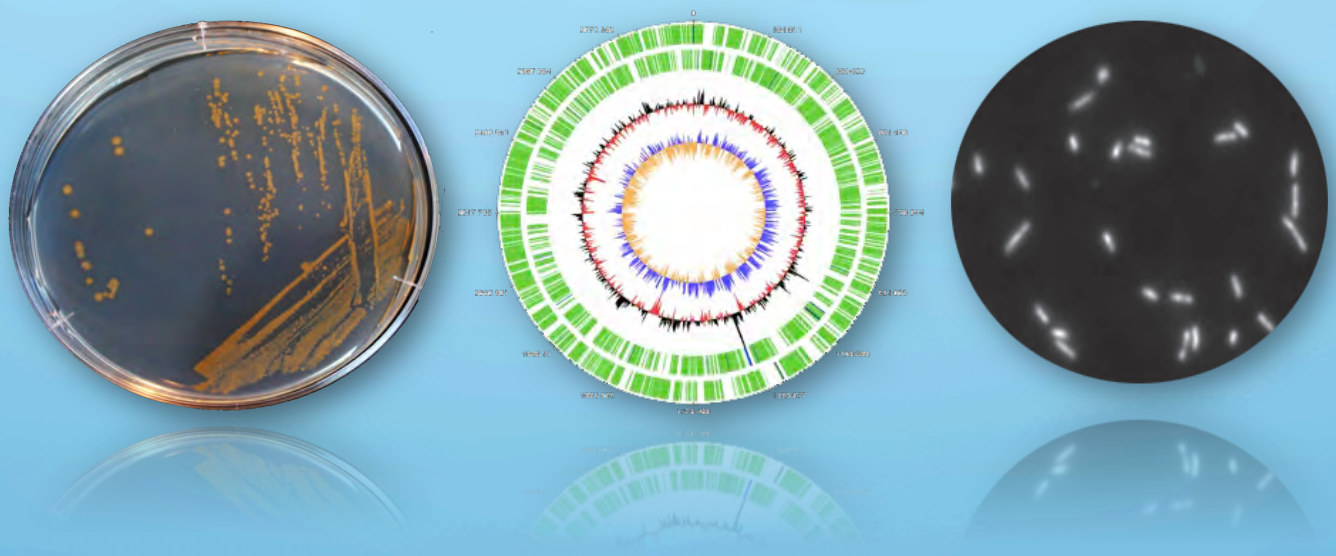
In vivo experiments

Culturing

As a complement to genomics, in **Chapter 4**, a considerable effort was conducted to develop a defined growth medium where the Bacteroidetes could grow in order to test questions regarding their physiological adaptations.

BIOLOG MT plates

These microplates were created for community analysis and microbial ecology studies (Garland and Mills, 1991; Insam, 1997), allowing the determination of the physiological characteristics of communities and to discern spatial and temporal changes. They consist on 96-well plates containing a carbon source and tetrazolium violet as a redox indicator of carbon source utilization. When bacteria oxidize the compound the indicator turns purple. We used them in **Chapter 5** as an attempt to detect the polymeric degradation activities of a microbial community.



Chapter 1

Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (Flavobacteria)

J. M. González, B. Fernández-Gómez, A. Fernàndez-Guerra, L. Gómez-Consarnau, O. Sánchez, M. Coll-Lladó, J. del Campo, L. Escudero, R. Rodríguez-Martínez, L. Alosno-Saéz, M. Latasa, I. Paulsen, O. Nedashkowskaya, I. Lekumberri, J. Pinhassi, and C. Pedrós-Alió

I contributed to this paper by manually annotating around 17% of the genes present in the genome of MED152 and subsequently I reconstructed several metabolic pathways: TCA cycle, pentose phosphate pathway, pyruvate metabolism, glycolisis, glyoxylate metabolism, Entner-Doudoroff pathway, and all amino acid biosynthesis pathways.

ABSTRACT

Analysis of marine cyanobacteria and proteobacteria genomes has provided a profound understanding of the life strategies of these organisms, their ecotype differentiation and metabolisms. However, a comparable analysis of the Bacteroidetes, the third major bacterioplankton group, is still lacking. In the present manuscript we report on the genome of *Polaribacter* sp. strain MED152. On the one hand, MED152 contains a substantial number of genes for attachment to surfaces or particles, gliding motility, and polymer degradation. This agrees with the currently assumed life strategy of marine Bacteroidetes. On the other hand, it contains the proteorhodopsin gene, together with a remarkable suite of genes to sense and respond to light, which may provide a survival advantage in the nutrient-poor, sun-lit ocean surface when in search of fresh particles to colonize. Furthermore, an increase in CO_2 fixation in the light suggests that the limited central metabolism is complemented by anaplerotic inorganic carbon fixation. This is mediated by a unique combination of membrane transporters and carboxylases. This suggests a dual life strategy that, if confirmed experimentally, would be notably different from what is known of the two other main bacterial groups (the autotrophic cyanobacteria and the heterotrophic proteobacteria) in the surface oceans. The *Polaribacter* genome provides new insights into the physiological capabilities of proteorhodopsin-containing bacteria. The genome will serve as a model to study the cellular and molecular processes in bacteria that express proteorhodopsin, their adaptation to the oceanic environment, and their role in carbon-cycling.

INTRODUCTION

Bacteroidetes are successful in the degradation of particulate organic matter in the ocean (1-2) and at least one molecular study showed them to be more abundant on particles that free-living in the water column (3). Many representatives have gliding motility and the capacity to degrade polymers, possibly allowing them to grow on detritus particles or algal cells using the polymeric substances as carbon and energy sources. Using microautoradiography combined with FISH, for example, Cottrell and Kirchman (4) showed that these bacteria are better adapted to the consumption of proteins over that of amino acids. Functional analysis of the genome of *Gramella forsetii* indicated that this marine Bacteroidetes has a substantial number of hydrolytic enzymes and a predicted preference for polymeric carbon sources (5). Although the diversity of Bacteroidetes is large, the adaptation to the degradation of polymeric substances seems a common theme. This trait contrasts that of another major group of marine bacteria, the proteobacteria: both alpha- and gammaproteobacteria seem to be better adapted to use monomers rather than polymers (4) and to a free-living existence in the water column. Therefore, the study of Bacteroidetes promises to reveal novel life strategies for successfully populating the surface ocean different from those of the proteobacteria whose complete genomes have been analyzed thus far (6-7).

Here we present the genome of *Polaribacter* sp. MED152. This genome was chosen for manual annotation and analysis for two reasons. In the first place it is representative of marine Bacteroidetes. Direct counts by fluorescence in situ hybridization repeatedly show Bacteroidetes to account for about 10-20% of the prokaryotes in seawater (8-9), most belonging to the flavobacteria (10-11). In 2004 there were a total of 864 16S rRNA gene sequences from marine Bacteroidetes in GenBank, of which 76 (9%) belonged to the genus *Polaribacter* (12). Out of the *Polaribacter* sequences 27 (36%) were most closely related to *Polaribacter dokdonensis*. Thus, *Polaribacter* is one of the major genera of Bacteroidetes found in the marine environment. In the second place, screening of the draft genome revealed the proteorhodopsin gene. The gene for this membrane protein was first found in DNA fragments directly obtained from seawater and functions as a light-driven H^+ pump in the ocean (13). Subsequent work has demonstrated a wide diversity and distribution of proteorhodopsin in the surface ocean bacterioplankton. *Escherichia coli* transformed with the proteorhodopsin gene can in fact use light energy for photophosphorylation (14) and cellular activities such as flagellar motion (15). Recently proteorhodopsin genes have been found in some cultured isolates (7, 16), a few belonging to the Bacteroidetes phylum (17). The presence of the proteorhodopsin gene in cultured bacteria opens the possibility to study the function of proteorhodopsin *in vivo*.

We have recently shown that *Dokdonia* sp. MED134 (a relative of *Polaribacter*) can use light energy gathered through proteorhodopsin to grow better in the light than in the dark (17). This is different from the light response of the alphaproteobacteria *Pelagibacter ubique* and gammaproteobacterium strain HTCC2207, neither one of which has been shown to grow better in the light despite the presence of a functional proteorhodopsin gene; refs. 10, 16). Accordingly, genome analysis of proteorhodopsin-containing flavobacteria opens a unique window to understand evolutionary adaptations to grow in a sun-lit environment. Our present genome analysis indicates that the strategy of *Polaribacter* sp. MED152 to grow in sea-water is different from that of other groups of abundant marine bacteria.

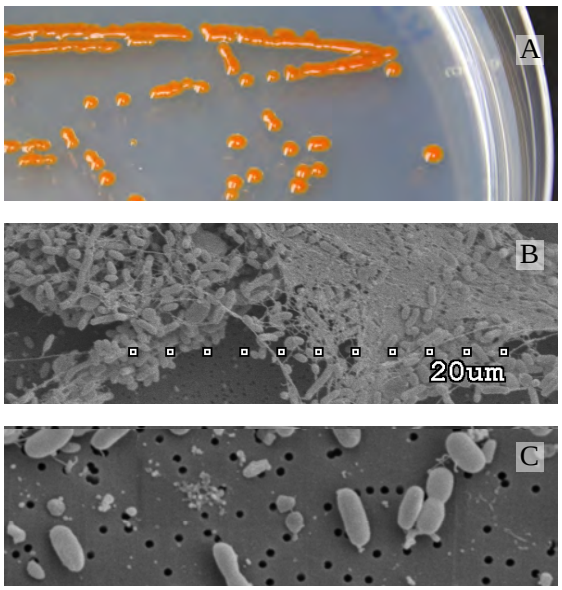


Figure 1. Images of MED152. A) Colonies on marine agar showing the characteristic orange color. **B)** SEM image of a typical aggregate showing abundant extracellular material. **C)** SEM image showing individual cells and extracellular material.

RESULTS AND DISCUSSION

Genome properties

MED152 forms bright orange colonies on agar plates and tends to aggregate into large flocks in liquid culture (Fig. 1). The genome contains 2,967,150 base pairs with 2692 predicted genes. This is a relatively small genome size for a marine bacterium. For example, 75% of the genomes in the Gordon and Betty Moore Foundation Marine Microbiology Initiative (total of 116 sequenced prokaryotes so far) have genomes larger than MED152 (with primarily SAR11 and *Prochlorococcus* genomes being smaller). Moreover, this is among the smallest genomes of Bacteroidetes isolates sequenced until now. The reduced genome size of MED152 is a consequence of a reduced number of protein coding genes and gene families compared to most other marine bacteria, in combination with a low number of paralogs in each family.

General genome features are presented in SI Table 1. Although not completely closed the genome sequence is on a single contig. G+C skew analysis indicates that the chromosome

is circular (data not shown). The largest protein families are peptidases (93 ORFs), glycosyl hydrolases (30 ORFs), TonB-dependent outer membrane channels (27 ORFs), response regulators (25 ORFs), glycosyl transferases (25 ORFs), ABC transporters (22 ORFs) and His kinases (21 ORFs).

Most peptides annotated as conserved hypothetical or as functional proteins are most similar to peptides in the closely related *Polaribacter irgensii* 23-P (1506 genes; based on BLASTP). Remaining peptides had homologues in other members of the Bacteroidetes phylum. It does not contain any IS element or peptides related to lysogenic phages, but five genes encode phage integrase family proteins that could allow for site-specific recombination.

Central metabolism

The metabolic pathways identified in the genome of MED152 are shown in SI Fig. 1-3, and SI Table 2. As expected in a bacterium living in surface waters, the genome of MED152 has adaptations for protection against stress from reactive oxygen species and repair of DNA damage (SI Tables 3-4). Consistent with its life in the ocean MED152 has a Na^+ rather than a H^+-translocating NADH:quinone oxidoreductase as a part of the respiratory chain to establish a Na^+ gradient for energy coupling (18).

There are two main characteristics of the central metabolism of MED152: one is logical and the other intriguing. The first one is logical considering the relatively small size of the genome: MED152 has a modest number of metabolic capabilities. It can only grow by aerobic respiration, although it does have a cbb3 type cytochrome oxidase mainly found in microaerophiles; it is unable to use fermentation or anaerobic respiration for energy conservation. Its genome also lacks the Entner-Doudouroff pathway; ammonia and sulfate are the single inorganic sources of nitrogen and sulfur respectively (SI Table 2). MED152 cannot use DMSP, or common products in algal exudates such as glycolate, taurine or polyamines, and it cannot oxidize inorganic sulfur compounds or carbon monoxide. We could only identify the metabolic pathways for the main cellular components (SI Table 2), but none of the rich variety of pathways found for example in *Silicibacter pomeroyi* (6).

The second characteristic is the remarkable number of genes potentially involved in anaplerotic metabolism (SI Table 2). First, MED152 has two types of putative bicarbonate importer genes similar to those found in cyanobacteria. The first one, *bicA* (MED152_09030, a SulP type, Na^+-dependent bicarbonate transporter) has a relatively low affinity for the substrate but a high flux rate (19); while the second one, *sbtA* (MED152_03855, also a Na^+-

dependent transporter) has high affinity for bicarbonate (20). These bicarbonate importer genes are complemented by a carbonic anhydrase gene that interconverts CO_2 and bicarbonate (Fig. 2).

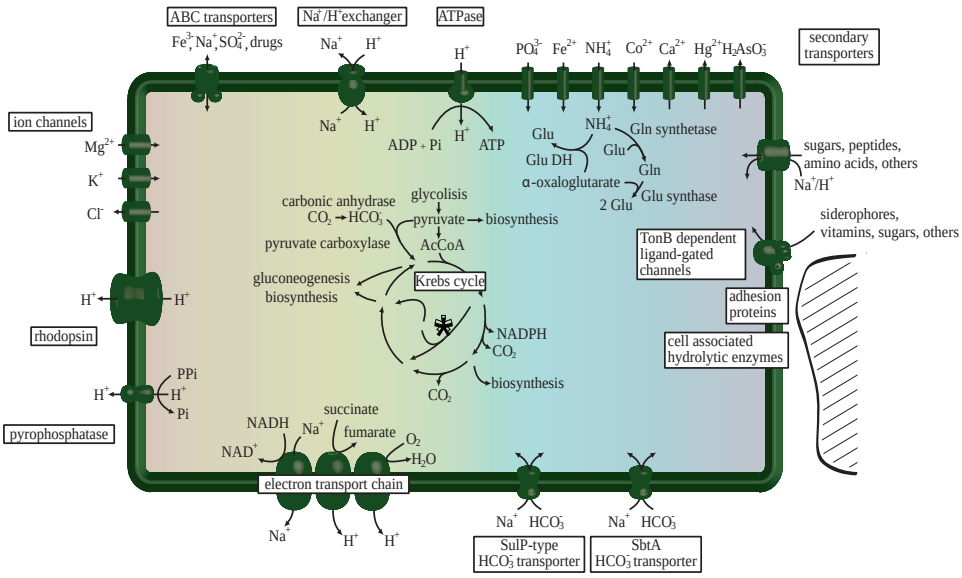


Figure 2. Diagram of a MED152 cell. Processes that generate a H^+ or Na^+ gradient are shown at the *Lower* and *Left*. Central metabolism pathways, as well as anaplerotic reactions, are indicated. Bicarbonate transporters are shown *Lower*.

Further, MED152 has two putative anaplerotic enzymes that utilize bicarbonate. These are PEP carboxylase (MED152_09950) and pyruvate carboxylase (MED152_04060) that generate oxaloacetate from bicarbonate and PEP, and from bicarbonate and pyruvate, respectively. Both enzymes are found in two other proteorhodopsin-containing Bacteroidetes, *P. irgensii* 23-P and *Dokdonia* sp. MED134. None of the remaining genome-sequenced marine proteobacteria with proteorhodopsin contain both pyruvate and PEP carboxylase genes (SI Table 5). Pyruvate carboxylase activity requires ATP, but PEP carboxylase does not. Since ATP generation in proteorhodopsin-containing bacteria could be independent of organic matter respiration in the presence of light, it may be speculated that some of this ATP could be used in biosynthesis through pyruvate carboxylase. In any case, the response of different proteorhodopsin-containing bacteria to light may differ depending on the genome context (14), in this case presence of specific transporters or carboxylases.

Carbonate concentrating mechanisms similar to those mentioned above play a critical role in CO_2-fixation pathways in autotrophic organisms, but could also be essential for anaplerotic CO_2 fixation in heterotrophs. MED152 does not have any of the conventional bicarbonate fixation pathways of autotrophs and it does not grow in a medium without organic matter. It is, therefore, a heterotroph. Growth of MED152 was measured in diluted

marine broth cultures exposed to light or darkness, to examine a potential benefit of harboring proteorhodopsin. In this relatively rich medium, growth was similar in the light and in the dark (Fig. 3, inset). In the related *Dokdonia* sp. MED134, growth in the light was better than in the dark only in medium with very low organic matter (17). We have not been successful at designing a simple poor medium where a light effect on growth could be shown at its

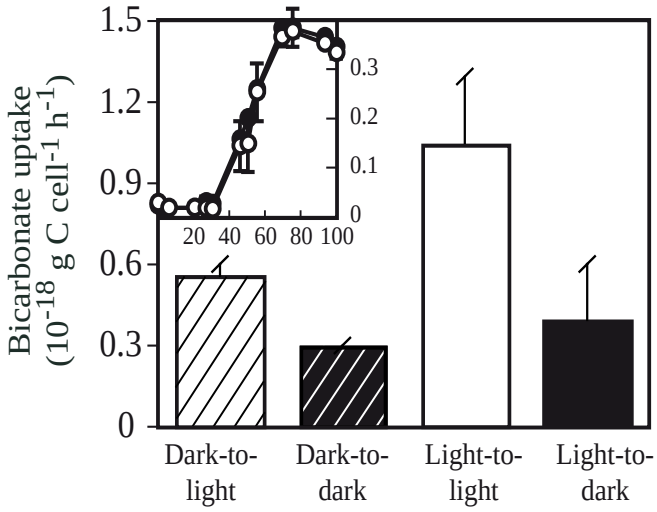


Figure 3. Bicarbonate uptake. Striped columns denote subsamples from cultures grown in the dark, incubated with $H^{14}CO_3^-$ in the light (Dark-to-light) or dark (Dark-to-dark). Solid columns indicate subsamples from cultures grown in the light, incubated with $H^{14}CO_3^-$ in the light (Light-to-light) or dark (Light-to-dark). (Inset) Growth of cultures in the dark (filled circles) and in the light (open circles) from where subsamples for bicarbonate uptake assays were collected at 50 h; the *x* axis shows time (hours) and the *y* axis optical density. Error bars denote SD for duplicates.

best if MED152 behaved exactly as *Dokdonia* sp. MED134. Nevertheless, experiments with radioactively labeled bicarbonate showed that, even though no differences were found in cell numbers, MED152 fixed more bicarbonate in the light than in the dark (Fig. 3). Notably, bicarbonate uptake rates were higher in the light irrespectively of light conditions prior to the uptake experiment. These results suggest that light-stimulated anaplerotic fixation of bicarbonate could play a significant role in the life strategy of MED152.

Proteorhodopsin phototrophy has been suggested to meet all energy requirements of the cell in the presence of light (15). Proteorhodopsin phototrophy allows the bacterium to invest reduced carbon substrates in biosynthesis instead of respiring them for ATP generation. This is advantageous in oligotrophic environments where labile organic matter is frequently in short supply. However, replacing respiration with biosynthesis represents a challenge to the cell, because it will result in an imbalance of TCA cycle intermediates necessary as precursors. Anaplerotic CO_2 fixation plays an important role in compensating for such

imbalances (SI Fig. 4). Thus, when intermediates of the TCA cycle are used for biosynthesis the TCA cycle needs to be replenished through the activity of bicarbonate fixation pathways. This could explain the presence of this type of bicarbonate concentrating mechanisms in MED152. Taken together, our findings suggest that light-stimulated anaplerotic inorganic carbon fixation could be a means allowing proteorhodopsin-containing flavobacteria to efficiently use organic matter for biosynthesis.

Light utilization and sensing

Despite its small genome, MED152 harbors an ample suite of genes involved in sensing and using light (SI Table 6) when compared to most other heterotrophic bacteria analyzed to date (21). The genome harbors the gene encoding proteorhodopsin and the genes for synthesis of the proteorhodopsin chromophore retinal (*crtEBIY*, *blh*). Moreover, the proteorhodopsin gene encodes a protein with conserved key amino acid positions required for light-driven H^+ pumping.

The synthesis of proteorhodopsin is expected to be regulated in response to light conditions. This has been shown in *Dokdonia* sp. MED134 (17). We therefore searched for genes potentially involved in such regulation. MED152 has a complex array of genes allowing it to respond to changes in the environment. The genome contains 21 two-component system sensor His kinases and 25 response regulators. In addition it contains 5 hybrid His kinase/ response regulators. Such sensor and signal transducing proteins likely provide means for controlling gene expression in response to environmental stimuli like light and nutrients.

A number of sensor membrane molecules contain domains that respond to light. Fig. 4 shows some of the remarkable architectures found in peptides with signaling domains. The PAS and GAF domains were each found in four of the two-component system sensor His kinases. PAS domains can respond to cellular energy levels, oxygen levels, redox potential and light (22), while GAF domains act as phototransducers. Both PAS and GAF domains are known to bind ligands and to interact with other proteins and they are common components of phytochromes. MED152, like proteorhodopsin-containing *Dokdonia* sp. MED134, contains a putative phytochrome known to detect red and far-red light (MED152_00100; SI Table 7) (23). This protein has a modular structure with PAS, GAF, phytochrome, HWE type His kinase and response regulator receiver domains. In cyanobacteria, similar phytochrome photoreceptors control gene expression through signal transduction pathways (24). This suggests that proteins with PAS and GAF domains, such as phytochromes, contribute to regulate the synthesis of proteins that respond to light in MED152.

Another remarkable set of genes in MED152 is that of the photolyases. Most of these are DNA repair flavoproteins that, in response to blue light, repair UV-induced DNA damage. The cryptochromes are a particular set of photolyase homologs, originally described in multicellular eukaryotes, where they function as blue light photoreceptors or regulators of circadian rhythm. Later, they have been described in prokaryotes as well (25). MED152 contained a genome region with three different DNA photolyase/cryptochrome genes (SI Fig. 5), while most marine bacteria had one or two genes only (SI Table 7). This region also contained a photolyase-related protein and *folE*, the latter possibly involved in the synthesis of the folate derived chromophore attached to the DNA photolyase/cryptochromes (26).

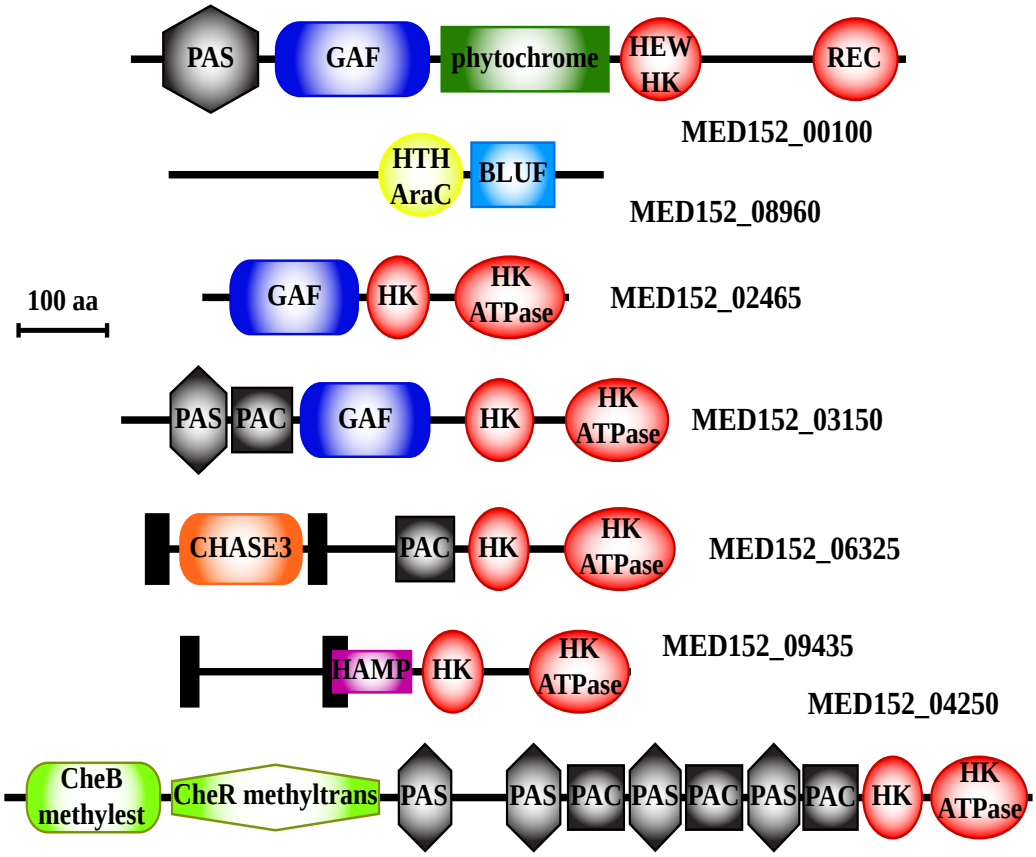


Figure 4. Domain architectures of selected peptides involved in signal transduction in MED152. MED152_00100 is a hybrid two-component His kinase sensor with both a HEW type His kinase (HEW HK) and a response regulator (REC). It contains a phytochrome domain that detects red and far-red light in a number of organisms. PAC, PAS and GAF domains are common components of phytochromes. HK is a His kinase domain with its cognate ATPase domain (HK ATPase). In MED152_08960 a BLUF domain (shown to sense blue light) is next to a DNA interacting, AraC type helix-turn-helix domain (HTC AraC). CHASE3 is an extracellular sensory domain and HAMP has been suggested to regulate signal transduction. CheB methylesterase and CheR methyltransferase are part of chemiotaxis signalling in bacteria. Solid bars represent transmembrane regions.

Phylogenetic analysis showed that each of the three DNA photolyase/cryptochrome genes in MED152 belonged to a different gene family (SI Fig. 6). The "class I CPD photolyase" and the "cryptochrome DASH" families are found in several Bacteroidetes genomes (SI Fig. 6, SI Table 7). Strikingly, however, proteorhodopsin-containing flavobacteria also have an additional DNA photolyase/cryptochrome belonging to the "animal cryptochrome and (6-4) photolyase" family; and they form a novel cluster within this family. Except for *P. ubique*, most marine bacteria with the proteorhodopsin gene have at least two types of DNA photolyase/cryptochrome (SI Table 7). If this correlation holds when more genomes are analyzed, some of these peptides are likely part of a light response mechanism that controls gene expression in marine bacteria.

Further, MED152_08960 contains one BLUF domain, which has been shown to sense blue light (Fig. 4). Similar domains are found in marine heterotrophic proteobacteria, aerobic anoxygenic photoheterotrophs and cyanobacteria (27). Interestingly the two proteorhodopsin-containing flavobacteria that contain a BLUF domain also contain a phytochrome (SI Table 7). None of these genes are found in other proteorhodopsin-containing bacteria.

MED152 is orange in color due to the presence of carotenoids (Fig. 1A). HPLC analysis showed that MED152 readily synthesizes three main carotenoids: β-carotene, zeaxanthin and a novel myxoxanthophyll-like carotenoid, and seven *crt* genes were identified. In addition to being precursors of retinal, carotenoids can function as accessory pigments in photosystems or protective antioxidants. Interestingly, most carotenoid synthesis genes in MED152 are localized in two gene clusters next to two different types of regulators, suggesting that they respond to different environmental stimuli. Thus, *crtZDA* are next to a two-component system response regulator of the LytR/AlgR type, while *crtYZBI* (responsible for the synthesis of β-carotene, the precursor of both retinal and the other carotenoids) is part of a different gene cluster next to a transcriptional regulator of the MerR family. *crt*Z is responsible for the synthesis of zeaxanthin from β-carotene and is present in both clusters. Similar gene content and organizations are found in other marine Bacteroidetes (e.g. *Dokdonia* sp. MED134, *Leeuwenhoekiella blandensis* MED217 and *P. irgensii* 23-P), regardless of the presence or absence of the proteorhodopsin gene. In summary, MED152 has a large range of genes that allow it to mediate, and possibly optimize, its reaction/response to light, a major environmental variable in the surface waters it was isolated from.

Adhesion and degradation of polymers

It has been proposed that marine flavobacteria are specialized in using proteins and other polymers for growth while proteobacteria would be specialized in using monomeric compounds (3-4, 28). Genome analysis of *G. forsetii* revealed an abundance of genes involved in attachment and a wide range of hydrolytic enzymes (5). These include 40 TonB dependent/ligand-gated channel genes (*susC*). In the related *Bacteroides thetaiotaomicron*, for example, the starch utilization system (Sus) operates in a way that the bacterium attaches to starch before starting its degradation (29-30). The TonB dependent/ligand-gated channel is an important component of the Sus complex. In these complexes there are hydrolytic enzymes located on the surface of the bacterium, but also outer membrane proteins that bind the polysaccharides to the bacterial surface. One such outer membrane protein is SusD (31). This life style requires these and other proteins to attach to the polymer of interest to degrade it, and to transport the monomers to the interior of the cell.

MED152 seems to have a similar system where the polymer degradation complex is associated to the cell surface. Twenty-seven genes were annotated as *susC* and five of these where next to their respective *susD* (SI Fig. 7). Other homologs of the eight-component Sus system in *B. thetaiotaomicron* were not found in the vicinity. However, the *susCD* were next to membrane protein genes with or without adhesion domains and also next to genes that encode enzymes that degrade polymers and transporters (SI Fig. 7). Although the TonB dependent/ligand-gated channels are better known for the iron uptake in Gram negative bacteria, they may be involved in the degradation and immediate uptake of the products of hydrolytic enzymes (32).

MED152 has a number of predicted periplasmic or extracellular peptidases and glycosyl hydrolases that is similar to Bacteroidetes that are well known for their capability to degrade polymeric compounds: *Flavobacterium johnsoniae*, *Flavobacterium psychrophilum*, *Cytophaga hutchinsonii* and *Bacteroides* spp. This suggests that proteins and complex carbohydrates are important carbon and nitrogen sources also for MED152. Indeed, in the laboratory, MED152 grows well on both peptone and starch. On the other hand, the number of transporters specific for free amino acids or oligopeptides was low, and very few dedicated sugar transporters were found (SI Table 8). Most notably, MED152 lacked sugar-specific ABC-transporters, while *Roseobacter* clade bacteria and marine gammaproteobacteria, in contrast, had up to 20 such transporters (SI Methods). In accordance with previous studies of substrate utilization in marine bacteria (4, 28), our data suggest that proteins and

polysaccharides, rather than monomers, are important carbon and nitrogen sources for at least certain marine flavobacteria.

Moreover, MED152 has a complete set of genes involved in gliding motility (15 genes), which could be beneficial in the exploration of solid surfaces. Suspended particles are particularly abundant during and following algal blooms, providing a potentially abundant source of food in the form of particulate organic matter. Flavobacteria are typically associated with the decay phase of phytoplankton blooms (1-2, 10). Further, up to 26 ORFs contained domains that are involved in attachment to surfaces (SI Table 9). Several of these ORFs were located within hydrolytic clusters that included *susCD* (SI Fig. 7).

MED152 has a large array of genes involved in the synthesis and export of extracellular polysaccharide material (e.g. 25 glycosyl transferases). Indeed, MED152 cells grown in liquid medium have a strong tendency to aggregate and form large visible flocks (Fig. 1B). Flavobacteria might also benefit from biofilm formation on particles, such as protection against grazing or against compounds that inhibit their growth and that are synthesized by competitors (33-34).

Sigma factors

As seen so far, MED152 has several sensors for light, possibly allowing regulation of gene expression under light and dark conditions. Further, it is reasonable to assume that the genome contains regulatory mechanisms for life attached to aggregates/particles compared to a free-living existence. As expected, the genome of MED152 encodes one RpoD type σ-70 transcription factor (MED152_11764) but has as many as 15 alternative σ-70 family of σ factors. Whereas *P. ubique* has only two and *E. coli* has six, the number of such transcription factors in other marine bacteria is as high as in MED152. Alternative sigma factors regulate the synthesis of particular sets of genes in response to different environmental stimuli. Thus, genes in the vicinity of the alternative σ-70 transcription factors could give clues as to the type of positively regulated genes and the type of signals MED152 might respond to. For example, MED152_02685 is next to a predicted transmembrane peptide, and both are next to a transcriptional regulator of the MerR type and genes involved in the synthesis of carotenoids and retinal. Both important features of *Polaribacter* metabolism, light utilization and polymer degradation, appear to be regulated by this mechanism since other σ-70 transcription factor genes are next to peptidase, glycosyl hydrolase, transporter and lipase genes.

Transporters

Membrane transporters play critical roles in making use of available nutrients as well as responding to variations in the environment. Genome analysis identified only 106 cell membrane transporters in MED152 (Table 1, SI Table 8). Our comparative analysis showed that this is a remarkably low number (see SI Methods for analysis details). For example, marine *Roseobacter* or gammaproteobacteria can harbor up to 330. The number of transporters in MED152 is only marginally higher than that in *P. ubique*, which has 88, despite the genome size of MED152 being more than double that of *P. ubique* (Table 1). A large number of transporters is typically found in versatile, and generalist bacteria. In contrast, a low number of transporters can be expected in specialized bacteria or bacteria living in a stable and predictable environment (35). Thus, it seems that MED152 is a relatively specialized bacterium when it comes to transporters, which is in accordance with its constrained cell metabolism and small genome size.

Table 1. Summary of transporters detected in the MED152 genome and selected genomes.

Transporters	PELAGIBACTER	MED152	ROSEOBACTER	*E. coli*
Total	88	106	210-330	~350
ATP-dependent	28	28	69-107	72
ABC superfamily	24	22	65-102	67
Ion channels	1	12	7-12	15
PTS	0	0	2-4	29
Secondary	57	62	112-156	233
DMT	20	6	17-37	16
MFS	7	10	14-25	70
MOP	3	7	4-7	8

PELAGIBACTER is *P. ubique* HTCC1062; ROSEOBACTER includes *Loktanella* sp. SKA53, *Roseobacter* sp. MED193, *Jannashia* sp. CCS1, *Roseovarius nubinhibens* ISM, *Sulfitobacte*r sp. TM1040, and *S. Pomeroyi* DSS-3; and *E. coli* is strain K12-MG1655.

CONCLUSION

The general features of the MED152 genome are consistent with life in the surface ocean, including both utilization of, and protection from, light and oxygen, and a significant number of Na^+ dependent proteins. We propose that the more specific features of the genome are related to the need of *Polaribacter* to alternate between two life styles. On the one hand, MED152 is very well equipped to attach to surfaces, glide in search for polymers and degrade them for carbon, nutrients and energy. On the other hand, once suitable polymeric substrates have been exhausted, MED152 needs to find new particles to colonize. This forces the bacterium to carry on a free-living existence in a carbon poor environment where it cannot move and is not prepared to take up most of the simple carbon compounds available. Thus, MED152 needs to somehow survive through the potentially long "traverse of the desert." MED152 seems to solve the problem by using proteorhodopsin to capture light energy, increase anaplerotic bicarbonate fixation, and use the few carbon compounds it can get hold of exclusively for biosynthesis. In addition, proteorhodopsin phototrophy could be useful in providing energy for the degradation of complex/recalcitrant organic matter. This strategy is completely different from that of proteobacteria and may be common to many of the other marine flavobacteria accounting for up to 30% of the prokaryotic cells in surface oceans. Analysis of the genome of MED152 has been extremely helpful in generating hypotheses about the life strategy for *Polaribacter* that can now be tested experimentally.

MATERIALS AND METHODS

Isolation of flavobacteria. Bacteria were isolated from Northwestern Mediterranean Sea surface water (0.5 m depth), collected 1 km off the Catalan coast at the Blanes Bay Microbial Observatory (41º 40' N, 2º 48' E; Spain). Strain MED152 was isolated on Marine Agar 2216 (Difco; Fig. 1A). 16S rRNA sequence analysis indicated that it belongs to the genus *Polaribacter* and its sequence is 99.6% similar to that of *P. dokdonensis* (SI Fig. 8). However, there are a few phenotypic differences between the two: unlike *P. dokdonensis*, MED152 is ß-galactosidase positive and is able to degrade gelatin.

Genomic sequencing and annotation. Whole-genome sequencing was carried out by the J. Craig Venter Institute through the Gordon and Betty Moore Foundation initiative in Marine Microbiology (www.moore.org). A Sanger/pyrosequencing hybrid method was used (36). Large (40 Kb) and small (4 Kb) insert random libraries were sequenced with an average success rate of 95% (large insert) and 93% (small insert) and an average high-

quality read length of 837 (large insert) and 860 (small insert) nucleotides. The completed genome sequence of MED152 contains 30,151 reads, achieving an average of 10-fold sequence coverage per base.

ORFs were predicted and autoannotated using GenDB (37). All predicted genes were used to query the TransportDB database (38) and matches were assigned to transporter families within the TransportDB database (http://www.membranetransport.org/). Further, a comparison to 19 genomes of marine bacteria in TransportDB was done (see SI Methods). All automatic annotations were curated manually. All predicted proteins from this genome were compared to selected genomes by using the basic local alignment search tool (BLAST). When necessary, analysis was carried out by using ad hoc Perl scripts.

Bacteria used in comparisons were members of the Bacteroidetes phylum in Genbank and genomes in the Gordon and Betty Moore Foundation Marine Microbiology Initiative. Gene content was also compared with available genomes of marine proteobacteria whose members are known to be abundant in sea water, such as the SAR11 and *Roseobacter* clades, as well as proteobacteria that contain the proteorhodopsin gene.

Bicarbonate uptake. MED152 was grown at 16°C in Marine Broth (Difco) diluted 1:8 with artificial seawater (35 practical salinity units, prepared from Sea Salts; Sigma) in light (180 µmol photons m^{-2} s^{-1}) or dark conditions. Bicarbonate uptake rates were determined by the radioactive carbon assimilation technique (39) after 50 hours incubation (during exponential growth; Fig. 3, inset): subsamples from each incubation condition were then placed in four 25 ml glass bottles (duplicates in both transparent and dark bottles), and incubation continued for 2 hours with 20 µl of $H^{14}CO_3^-$ (3 µCi). For details see SI Methods.

Pigment analysis. Pigments were identified and quantified using HPLC analysis as described in SI Methods.

ACKNOWLEDGEMENTS

We thank Folker Meyer, Jomuna Veronica Choudhuri, Daniela Bartels, Andreas Wilke and Tobias Paczian, then at the University of Bielefeld, Germany, for their help with GenDB and José Manuel Fortuño of ICM for help with electron microscopy.
The genome was sequenced through the Marine Microbial Initiative of the Gordon and Betty Moore Foundation. This work was supported by the European Union (contract EU-FP6-505403 NoE Marine Genomics Europe), the Spanish Ministry of Education and Science (grant CTM2004-02586/MAR) and the Swedish Research Council (grant 621-2003-2692).

REFERENCES

1. Pinhassi J, Sala MM, Havskum H, Peters F, Guadayol Ò, Malits A, Marrasé C (2004) Changes in bacterioplankton composition under different phytoplankton regimens. *Appl. Environ. Microbiol.* 70:6753-6766.
2. Riemann L, Steward GF, Azam F (2000) Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. *Appl. Environ. Microbiol.* 66:578-587.
3. DeLong E, Franks DG, Alldredge AL (1993) Phylogenetic diversity of aggregate-attached versus free-living marine bacterial assemblages. *Limnol. Oceanog.* 38:924-934.
4. Cottrell MT, Kirchman DL (2000) Natural assemblages of marine Proteobacteria and members of the *Cytophaga-Flavobacter* cluster consuming low- and high-molecular-weight dissolved organic matter. *Appl. Environ. Microbiol.* 66:1692-1697.
5. Bauer MM, Kube H, Teeling M, Richter, Lombardot T *et al.* (2006) Whole genome analysis of the marine *Bacteroidetes* '*Gramella forsetii*' reveals adaptations to degradation of polymeric organic matter. *Environ. Microbiol.* 8:2201-2213.
6. Moran MA, Buchan A, González JM, Heidelberg JF, Whitman WB *et al.* (2004) Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* 432:910-913.
7. Giovannoni SJ, Bibbs L, Cho JC, Stapels MD, Desiderio R *et al.* (2005) Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* 438:82-85.
8. Cottrell MT, Kirchman DL (2000) Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence in situ hybridization. *Appl. Environ. Microbiol.* 66:5116-5122.
9. Glöckner F-O, Fuchs BM, Amann RI (1999) Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence in situ hybridization. *Appl. Environ. Microbiol.* 65:3721-3726.
10. Abell GCJ, Bowman JP (2005) Ecological and biogeographic relationships of class Flavobacteria in the Southern Ocean. *FEMS Microbiol. Ecol.* 51:265-277.
11. Alonso C, Warnecke F, Amann RI, Pernthaler J (2007) High local and global diversity of Flavobacteria in marine plankton. *Environ. Microbiol.* 9:1253-1266.
12. Pommier T, Pinhassi J, Hagström Å (2005) Biogeographic analysis of ribosomal RNA clusters from marine bacterioplankton. *Aquat. Microb. Ecol.* 41:79-898.
13. Béjà O, Aravind L, Koonin EV, Suzuki MT, Hadd A *et al.* (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* 289:1902-1906.
14. Martinez A, Bradley AS, Waldbauer JR, Summons RE, DeLong EF (2007) Proteorhodopsin photosystem gene expression enables photophosphorylation in a heterologous host. *Proc. Natl. Acad. Sci. USA* 104: 5590-5595.

15. Walter JM, Greenfield D, Bustamante C, Liphardt J (2007) Light-powering *Escherichia coli* with proteorhodopsin. *Proc. Natl. Acad. Sci. USA* 104:2408-2412.
16. Stingl U, Desiderio RA, Cho JC, Vergin KL, Giovannoni SJ (2007) The SAR92 clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. *Appl. Environ. Microbiol.* 73:2290-2296.
17. Gómez-Consarnau L, González JM, Coll-Lladó M, Gourdon P, Pascher T *et al.* (2007) Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. *Nature* 445:210-213.
18. Unemoto T, Hayashi M (1993) Na^+-translocating NADH-quinone reductase of marine and halophilic bacteria. *J. Bioenerg. Biomembr.* 25:385-391.
19. Price GD, Woodger FJ, Badger MR, Howitt SM, Tucker L (2004) Identification of a SulP-type bicarbonate transporter in marine cyanobacteria. *Proc. Natl. Acad. Sci. USA* 101:18228-18233.
20. Shibata M, Katoh H, Sonoda M, Ohkawa H, Shimoyama M *et al.* (2002) Genes essential to sodium-dependent bicarbonate transport in cyanobacteria. *J. Biol. Chem.* 277:18658-18664.
21. Van der Horst MA, Key J, Hellingwerf KJ (2007) Photosensing in chemotrophic, non-phototrophic bacteria: let there be light sensing too. *Trends Microbiol.* 15:554-562.
22. Taylor BL, Zhulin IB (1999) PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol. Mol. Biol. Rev.* 63:479-506.
23. Fankhauser C (2001) The phytochromes, a family of red/far-red absorbing photoreceptors. *J. Biol. Chem.* 276:11453-11456.
24. Ashby MK, Houmard J (2006) Cyanobacterial two-component proteins: structure, diversity, distribution, and evolution. *Microbiol. Mol. Biol. Rev.* 70:472-509.
25. Hitomi K, Okamoto K, Daiyasu H, Miyashita H, Iwai S *et al.* (2000) Bacterial cryptochrome and photolyase: characterization of two photolyase-like genes of *Synechocystis* sp. PCC6803. *Nucleic Acids Res.* 28:2353-2362.
26. Kanai S, Kikuno R, Toh H, Ryo H, Todo T (1997) Molecular evolution of the photolyase-blue-light photoreceptor family. *J. Mol. Evol.* 45:535-548.
27. Gomelsky M, Klug G (2002) BLUF: a novel FAD-binding domain involved in sensory transduction in microorganisms. *Trends Biochem. Sci.* 27:497-500.
28. Kirchman DL (2002) The ecology of *Cytophaga-Flavobacteria* in aquatic environments. *FEMS Microbiol. Ecol.* 39:91-100.
29. Anderson KL, Salyers AA (1989) Biochemical evidence that starch breakdown by *Bacteroides thetaiotaomicron* involves outer membrane starch binding sites and periplasmic starch-degrading enzymes. *J. Bacteriol.* 171:3192-3198.
30. Anderson KL, Salyers AA (1989) Genetic evidence that outer membrane binding of

starch is required for starch utilization by *Bacteroides thetaiotaomicron*. *J. Bacteriol.* 171:3199-3204.

31. Reeves AR, Wang G-R, Salyers AA (1997) Characterization of four outer membrane proteins that play a role in utilization of starch by *Bacteroides thetaiotaomicron*. *J. Bacteriol.* 179:643-649.
32. Blanvillain S, Meyer D, Boulanger A, Lautier M, Guynet C *et al.* (2007) Plant carbohydrate scavenging through TonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. *PLoS ONE* 2:e224.
33. Long RA, Azam F (2001) Antagonistic interactions among marine pelagic bacteria. *Appl. Environ. Microbiol.* 67:4975-4983.
34. Grossart HP, Schlingloff A, Bernhard M, Simon M, Brinkhoff T (2004) Antagonistic activity of bacteria isolated from organic aggregates of the German Wadden Sea. *FEMS Microbiol. Ecol.* 47:387-396.
35. Ren Q, Paulsen IT (2005) Comparative analyses of fundamental differences in membrane transport capabilities in Prokaryotes and Eukaryotes. *PLoS Comput. Biol.* 1:e27.
36. Goldberg MD, Johnson J, Busam D, Feldblyum T, Ferriera S *et al*. (2006) A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. *Proc. Natl. Acad. Sci. USA* 103:11240-1124.
37. Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T *et al.* (2003) GenDB – an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.* 31:2187-2195.
38. Ren Q, Kang KH, Paulsen IT (2004) TransportDB: a relational database of cellular membrane transport systems. *Nucleic Acids Res.* 32:D284-D288.
39. Parsons TR, Maita Y, Lalli CM (1984) A manual of chemical and biological methods for seawater analysis. Pergamon Press, Oxford, UK.

SUPPORTING INFORMATION

SI METHODS

Transporters. The low number of transporters was apparent in most of the main transporters families. Thus, together with *Pelagibacter ubique*, MED152 has the lowest number of ATP-dependent transporters (a majority belonging to the ATP-binding Cassette [ABC] superfamily) of the marine bacteria sequenced so far (Table 1). The same holds true for the secondary transporters, which account for 50 to 70% of the transporters in the marine bacterial genomes sequenced until now. MED152 has only a total of 62 secondary transporters – similar to the number found in *Pelagibacter ubique* – while other bacteria may have from 100-200. In particular, the number of transporters in the drug/metabolite transporter (DMT) superfamily and in the major facilitator superfamily (MFS; used for transport of a diverse set of "small solutes in response to chemiosmotic ion gratients"; Pao *et al.*, 1998) transporters was low, while the number of multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) flippase superfamily transporters was similar to that in other marine bacteria (despite its smaller genome).

Like *Pelagibacter ubique*, MED152 (and other marine flavobacteria) lacks phosphotransferase system (PTS) genes for active transport of sugars. PTS genes are present in all major bacterial groups except cyanobacteria and epsilonproteobacteria (Barabote and Saier, 2005). The lack thereof is consistent with the aerobic life style of MED152 (PTS are particularly important for sugar uptake under anaerobic conditions).

Comparative analysis of transporters. Transporter families in MED152 were compared to the following marine genomes in the TransportDB database (details on transporter families and substrate specificity of particular transporters in specific bacteria are available at http://www.membranetransport.org/): alphaproteobacteria *Bradyrhizobium japonicum* USDA 110, "*Candidatus* Pelagibacter ubique" HTCC1062, *Caulobacter crescentus* CB15, *Hyphomonas neptunium* ATCC 15444, *Jannaschia* sp. CCS1, *Loktanella vestfoldensis* SKA53, *Maricaulis maris* MCS10, *Roseobacter* sp. MED193, *Roseobacter* sp. TM1040, *Roseovarius nubinhibens* ISM, and *Silicibacter pomeroyi* DSS-3; gammaproteobacteria *Aeromonas hydrophila* subsp. *hydrophila* ATCC 7966, *E. coli* K12-MG1655, *Marinomonas* sp. MED121, *Photobacterium profundum* SS9, *Pseudoalteromonas haloplanktis* TAC125, *Pseudomonas aeruginosa* PAO1, *Shewanella oneidensis* MR-1, and *Vibrio* sp. MED222; Bacteroidetes/Chlorobi phylum *Bacteroides fragilis* NCTC9343, *Bacteroides fragilis* YCH46, *Bacteroides thetaiotaomicron* VPI-5482, *Chlorobium tepidum* TLS, *Dokdonia* sp. MED134,

Leeuwenhoekiella blandensis MED217, and *Polaribacter* sp. MED152; and Firmicutes *Bacillus subtilis* 168.

Bicarbonate uptake. For the bicarbonate uptake experiment, MED152 was grown at 16°C in Marine Broth (Difco) diluted 1:8 with artificial seawater (35 practical salinity units, prepared from Sea Salts; Sigma) in light (180 µmol photons m^{-2} s^{-1}) or dark conditions (duplicate 200 ml cultures in each condition). Cultures were maintained without shaking to reduce aggregation and flock formation. After 50 hours incubation (during exponential growth), four subsamples from one culture of each incubation condition (light and dark) were transferred to 25 ml glass bottles. From each original incubation condition, two subsamples (treated as duplicates) were placed in transparent bottles and two subsamples were placed in dark bottles (bottles covered with aluminum foil and black plastic). All bottles were then incubated under white light (180 µmol photons m^{-2} s^{-1}; no light entered the dark bottles) for 2 hours with 20 µl of $H^{14}CO_3^-$ (3 µCi; DHI, Denmark). Controls from each treatment were treated with 10% trichloroacetic acid (final concentration). After incubation, 3 ml aliquots from each replicate bottle were filtered through 0.2 µm pore size filters (25 mm diameter, Supor-200, Pall), and the filters were exposed to HCl 0.7 M fumes for 2 hours. Finally, the filters were placed in vials with 10 ml scintillation cocktail (Perkin Elmer) and kept at least 24 hours in the dark before counting. Bicarbonate uptake rates were calculated according to the standard radioactive carbon assimilation technique procedure (Parsons *et al*., 1984). At the time of the experiment, bacterial abundance was approximately 5.1 $\times 10^8$ cells ml^{-1}, as determined by epifluorescence microscopy of SYBR Gold stained cells.

Pigment analysis. MED152 was grown in Marine Broth 2216 (Difco) and filtered onto Whatman GF/F filters. Pigments were extracted by placing the filters in 3 mL of 90% acetone (with 0.01% of butylated hydroxytoluene to prevent chlorophyll allomerization) and vortexing them vigorously for 45 s. After 24 h at -20ºC, samples were sonicated for 1 min and vortexed again for 45 s. The extracts were cleared by filtration through Poretics 0.8 µm polycarbonate filters. For pigment chromatography, 150 µL of a mixture of 0.5 mL extract plus 0.1 mL H_2O was injected into a Thermo HPLC system and run under the conditions described in Latasa *et al.* (2001). Standards of myxoxanthophyll, zeaxanthin and β-carotene (DHI, Denmark) were used for pigment identification and quantification.

1. Pao SS, Paulsen IT, Saier MH (1998) Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* 62:1-34.

2. Barabote RD, Saier MH (2005) Comparative genomic analyses of the bacterial phosphotransferase system. *Microbiol. Mol. Biol. Rev.* 69:608-634.

3. Parsons TR, Maita Y, Lalli CM (1984) A manual of chemical and biological methods for seawater analysis. Pergamon Press, Oxford, UK.

4. Latasa M, van Lenning K, Garrido JL, Scharek, R, Estrada M *et al.* (2001) Losses of chlorophylls and carotenoids in aqueous acetone and methanol extracts prepared for RPHPLC analysis of pigments. *Chromatographia* 53: 385-391.

5. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673-4680.

6. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17:540-552.

7. Felsenstein J (1989) PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics* 5:164–166.

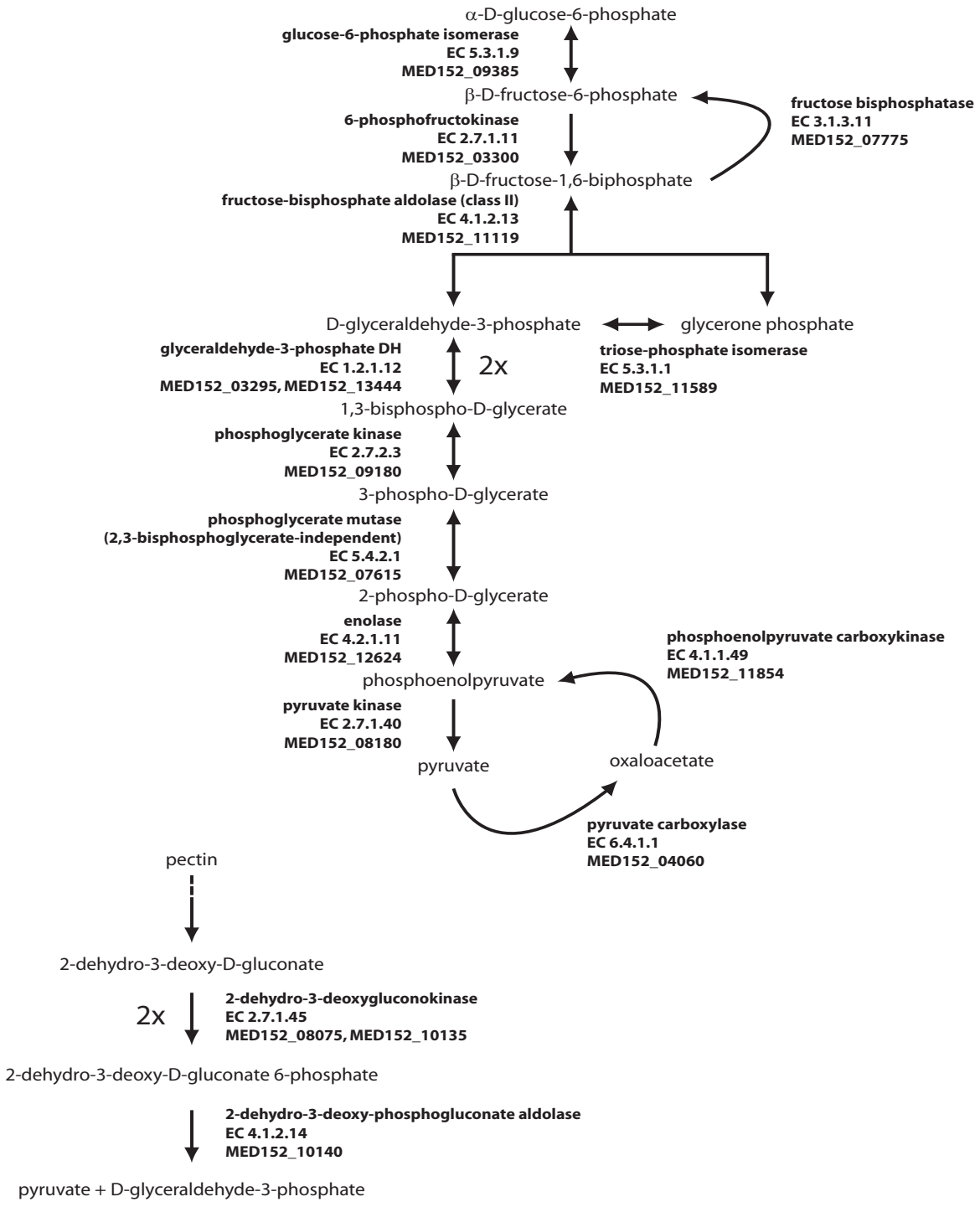


Figure S1. Glycolisis/gluconeogenesis.

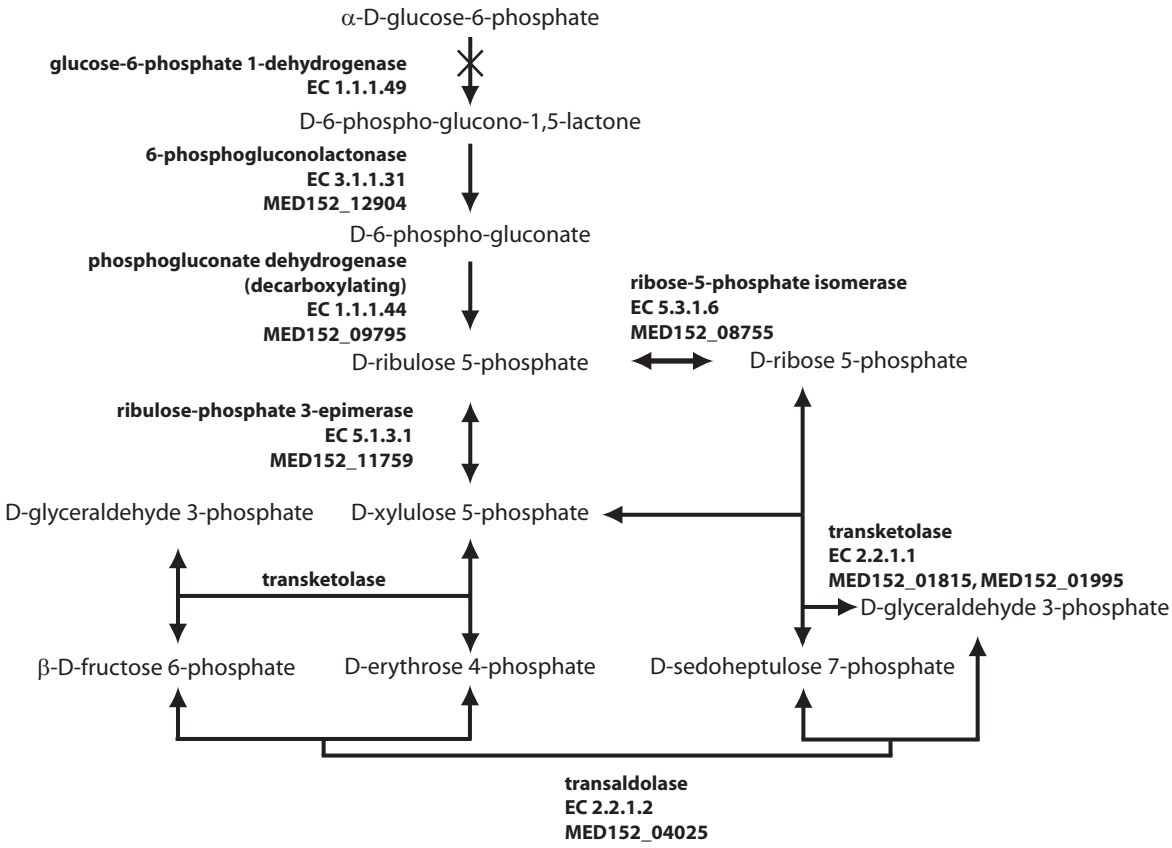


Figure S2. Pentose phosphate pathway.

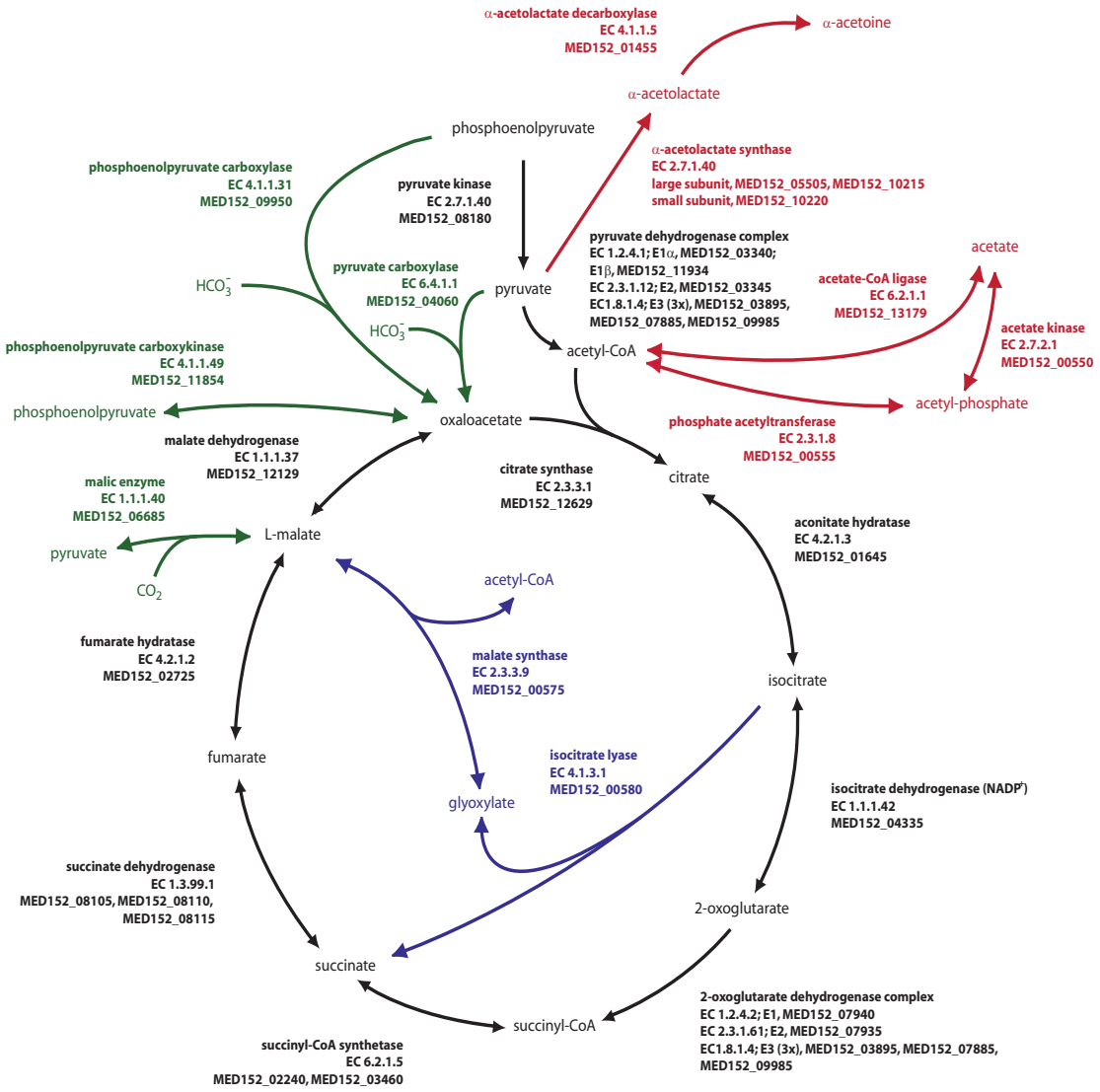


Figure S3. TCA cycle and glyoxylate shunt.

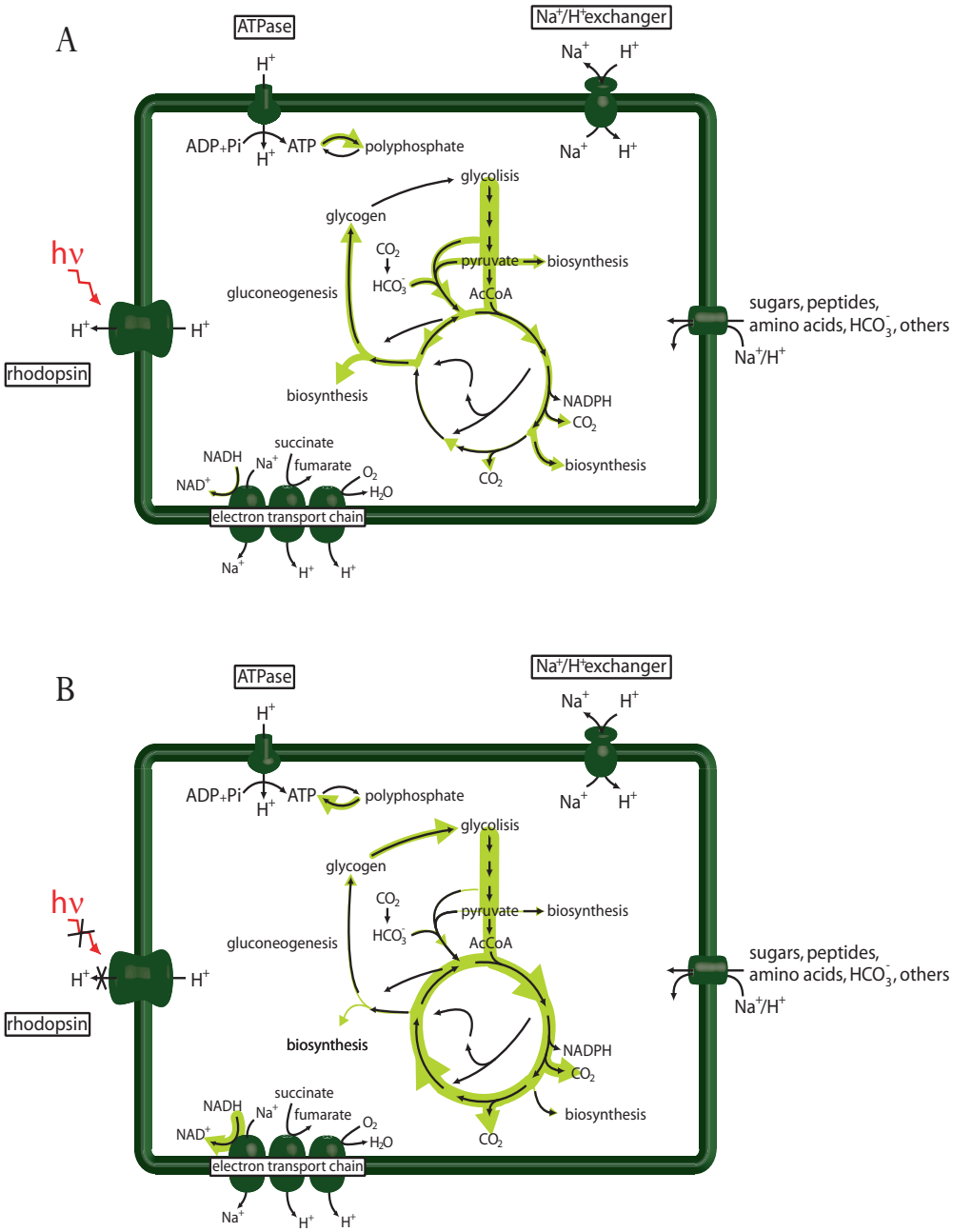


Figure S4. Proposed carbon flow in MED152 in the presence and absence of light. When the cell is exposed to light **(A)**, the prevailing role of the main metabolic pathways is biosynthetic. In the dark **(B)**, the cell must respire more organic carbon than in the light to meet its energy requirements.

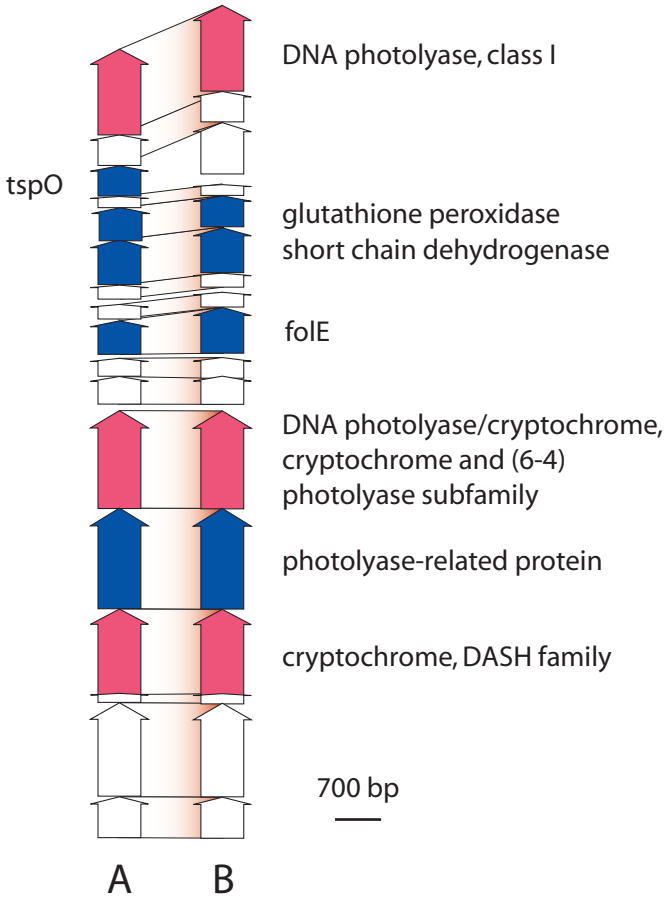


Figure S5. *Polaribacter irgensii* 23-P **(A)** and *Polaribacter* sp. MED152 **(B)** gene clusters that contain the three copies of the DNA photolyase/cryptochrome genes (in red). Genes in white are conserved hypothetical. TspO is a Trp-rich sensory protein that acts as a transcriptional regulator and is only found in the genome of *Polaribacter irgensii* 23-P but not in MED152.

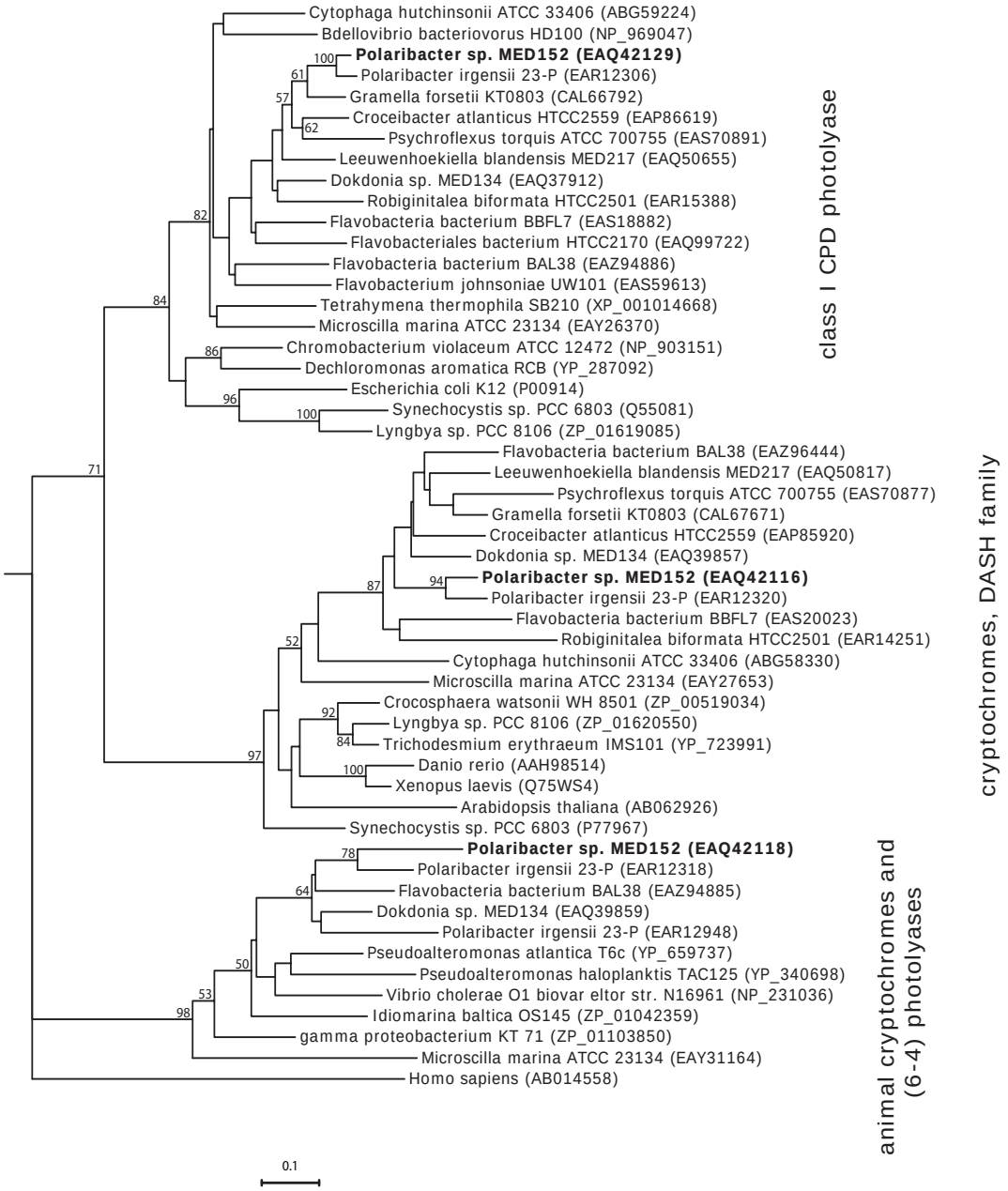


Figure S6. Evolutionary relationships of cryptochrome/photolyase protein family from the marine Bacteroidetes and representatives from other organisms. Subfamilies are indicated on the right. A multiple alignment was constructed with the software package CLUSTAL W 1.74 (Thompson *et al.*, 1994). The alignment was edited with Gblocks (Version 0.91b) to identify conserved regions (Castresana, 2000) with a minimum block of ten and without gaps. The tree was constructed based on a Kimura's distance matrix and the Neighbour-Joining method using the PHYLIP package (Version 3.2) (Felsenstein, 1989). The statistical significance of the tree topology was evaluated by bootstrap analysis with 1000 iterative constructions of the neighbour-joining tree. The numbers at the nodes are bootstrap values higher than 50%. The scale bar represents the Kimura distance.

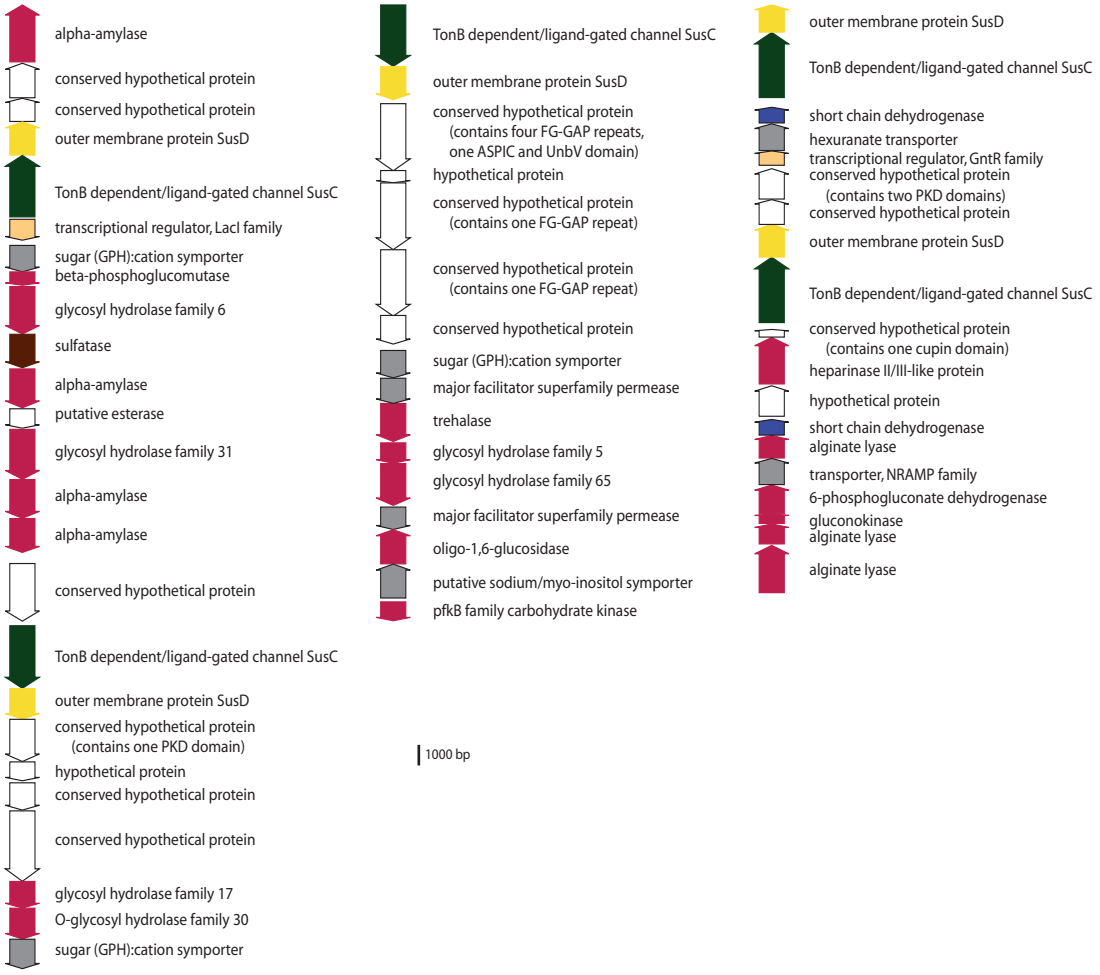


Figure S7. Clusters of genes putatively involved in the attachment and degradation of polymeric compounds. Only those that contain the tandem TonB dependent/ligand-gated channel genes and *susD* homologs are represented.

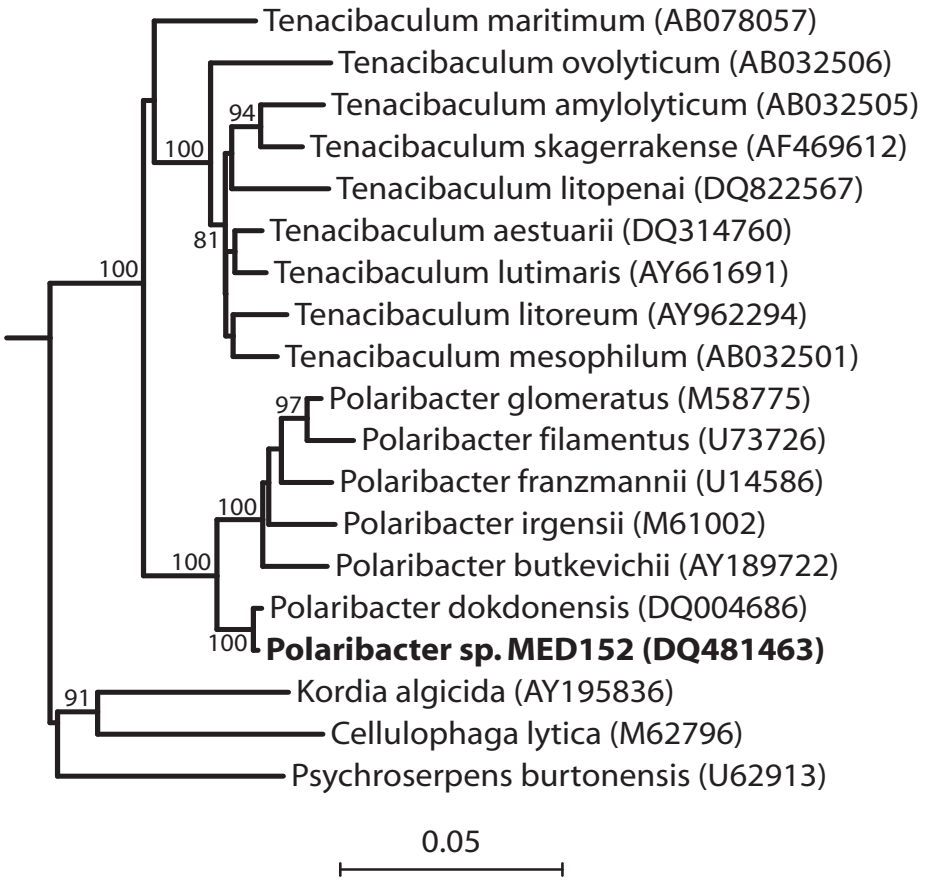


Figure S8. Neighbour-joining phylogenetic tree based on 16S rRNA gene sequences of MED152 and closely related type strains for each species. A multiple alignment was constructed with the software package CLUSTAL W (Version 1.83; Thompson *et al.*, 1994). The alignment was edited with Gblocks (Version 0.91b) to identify conserved regions (Castresana, 2000) with a minimum block of ten and without gaps. The tree was constructed based on a Jukes-Cantor distance matrix and the Neighbour-Joining method using the PHYLIP package (Version 3.2) (Felsenstein, 1989). The sequence of *Cytophaga hutchinsonii* (M58768) served as the outgroup. Bootstrap values greater than 70% confidence are shown at branching points (percentage of 1000 resamplings). GenBank accession numbers are given in parentheses. The scale bar represents Jukes-Cantor distance.

Supplementary Table 1. Summary of MED152 genome.

Parameter	Value
No. of bp	2,967,150
GC content (%)	30.61
Coding density (%)	92.6
No. of predicted protein coding genes	2,646
No. of predicted proteins with putative function	1,750
No. of conserved hypothetical proteins	593
No. of unknown proteins unique to MED152	303
No. of tRNA's	37
No. of rrn	3

Table S2. Key enzymes and metabolic pathways identified in MED152

Pathway/enzyme (gene copy when greater than one)	ORFs
Main pathways	
Pyruvate carboxylase. Anaplerotic pathway	MED152_04060
PEP carboxylase. Anaplerotic pathway	MED152_09950
Glyoxylate bypass. Anaplerotic pathway	MED152_00580 (aceA), MED152_00575 (aceB)
Malic enzyme. Anaplerotic pathway	MED152_06685
Phosphate acetyltransferase. Acetate metabolism	MED152_00555
Acetate kinase. Acetate metabolism	MED152_00550
Acetate CoA ligase. Acetate metabolism	MED152_13179
Acyl-CoA dehydrogenase. β-Oxidation (8)	MED152_01705, MED152_04870, MED152_05285, MED152_05295, MED152_05785, MED152_07525, MED152_07535, MED152_10560
Enoyl-CoA hydratase. β-Oxidation (4)	MED152_05285, MED152_06790, MED152_09440, MED152_11649
Electron transfer protein. β-Oxidation	MED152_11949 (etfA), MED152_11944 (etfB)
Glycerol kinase. Lipid degradation	MED152_08680
Glycerol-3-phosphate dehydrogenase (NAD(P)+)	MED152_05430, MED152_08715
FAD-linked glycerol-3-phosphate dehydrogenase.	MED 152_08690
Glycogen biosynthesis	MED152_05855 (glgA), MED152_05860 (glgC), MED152_05865, MED152_05870, MED152_05875 (glgB), MED152_05880, MED152_05885
Electron transport chain	
Na^+-translocating NADH:quinone oxidoreductase	MED152_11704 (nqrA), MED152_11709 (nqrB), MED152_11714 (nqrC), MED152_11719 (nqrD), MED152_11724 (nqrE), MED152_11729 (nqrF)
Type II NADH dehydrogenase (FAD-dependent)	MED152_12814
Succinate dehydrogenase	MED152_08110 (sdhA), MED152_08115 (sdhB), MED152_08105 (sdhC)
Cytochrome c oxidase	MED152_10275 (coxM), MED152_10280 (coxN), MED152_06775 (coxO), MED152_06770 (coxP)
Cytochrome oxidase, cbb3 type	MED152_03190 (ccoNO), MED152_03195 (ccoS)
Cytochrome c class I	MED152_10265
Cytochrome c	MED152_10245
Cytochrome c assembly protein	MED152_05520
Cytochrome oxidase assembly protein	MED152_06660
Cytochrome oxidase assembly factor	MED152_06780
60 KD inner membrane protein OxaA homolog	MED152_13299
SenC (3)	MED152_03785, MED152_06550, MED152_06755
Nitrogen assimilation	
Glutamate synthase	MED152_05815 (gltB), MED152_05820 (gltD)
Ammonium channel (amtB) (2)	MED152_05800, MED152_05810
Nitrogen regulatory protein P-II	MED152_05805
Glutamine synthetase, type II	MED152_05925
Glutamine synthetase, type III	MED152_05920
Glutamate dehydrogenase	MED152_08660
Phosphate metabolism	
Phosphate permease	MED152_09510
Polyphosphate kinase (3)	MED152_00300, MED152_00305, MED152_05065
Exopolyphosphatase	MED152_04590
H^+-translocating pyrophosphatase	MED152_11924
Soluble pyrophosphatase	MED152_11929
Sulfur assimilation	
APS/PAPS pathway	MED152_06170 (cysD), MED152_06175 (cysH), MED152_06165 (cysN), MED152_06160 (cysI), MED152_09765 (cysJ), MED152_00895 (cysK), MED152_06130 (cysM), MED152_06135 (cysE)
Sulphate permease (3)	MED152_09030, MED152_10290, MED152_12914
Iron assimilation	
Fe^{3+}-binding periplasmic protein (2)	MED152_01595, MED152_11574
ABC-type Fe^{3+}-siderophore transporter (2)	MED152_00690, MED152_00695
Mn^{2+}/Fe^{2+} transporter, NRAMP family (2)	MED152_03840, MED152_11859
Ferritin-like protein (2)	MED152_01695, MED152_03075
ATP-dependent Fe^{2+} transport system FeoAB	MED152_06540, MED152_06545
Fur (2)	MED152_00170, MED152_01940
Repressor	MED152_10760
Bicarbonate uptake	
SulP type transporter (BicA)	MED152_09030
SbtA	MED152_03855
Carbonic anhydrase	MED152_09035

The following additional pathways were complete and are not shown on the table: glycolysis, gluconeogenesis, TCA cycle, pentose phosphate pathway, purine biosynthesis and salvage, pyrimidine biosynthesis and salvage, thimidylate biosynthesis, amino acid metabolism, fatty acid biosynthesis, NAD biosynthesis, riboflavin and FAD biosynthesis, siroheme biosynthesis, quinone biosynthesis, H^+-ATPase, pyridoxal phosphate biosynthesis, pantothenate and CoA biosynthesis. Complete cobalamine, biotin and thiamine biosynthetic pathways were not found. Entner–Doudoroff pathway is incomplete.

Table S3. Key stress response-related genes in MED152

Pathway/protein (copy number when greater than one)	Type of stress	ORFs
Glycine betaine biosynthesis	Osmotic	MED152_05190 (betA), MED152_04445 (betB)
OsmC (2)	Osmotic	MED152_04480, MED152_07905
Large Conductance Mechanosensitive Ion Channel (MscL) family (2)	Osmotic	MED152_07755*, MED152_08290*
Small Conductance Mechanosensitive Ion Channel (MscS) family (6)	Osmotic	MED152_03080*, MED152_05265*, MED152_06730*, MED152_10700*, MED152_11349*, MED152_12279*
Voltage-Gated K^+ Channel superfamily (2)	Osmotic	MED152_01430*, MED152_04890*
K^+ uptake protein TrkH	Osmotic	MED152_12369*
Monovalent Cation:Proton Antiporter-2 (CPA2) family K^+ transporter	Osmotic	MED152_02125*
Cl^- Channel (ClC) family	Osmotic	MED152_07855*
ATP-Binding Cassette (ABC) type Na^+ transporter (3)	Osmotic	MED152_05270*, MED152_05275*, MED152_13214*
Phage shock protein C (PspC)	Osmotic/cryoprotection	MED152_03405
Fatty acid desaturase (3)	Cryoprotection	MED152_06405, MED152_06505, MED152_09405
NhaC Na^+:H^+ Antiporter (NhaC) family (2)	Internal pH regulation/Na^+ extrusion	MED152_06020*, MED152_08675*
NhaD Na^+:H^+ Antiporter (NhaD) family	Internal pH regulation/Na^+ extrusion	MED152_05770*
Monovalent Na^+/H^+ Antiporter-1 (CPA1) family (2)	Internal pH regulation/Na^+ extrusion	MED152_02960*, MED152_03050*
Sterol desaturase (3)	Cryoprotection	MED152_02375, MED152_07040, MED152_10110
DEAD box RNA helicase (12)	Cryoprotection	MED152_00240, MED152_01675, MED152_01850, MED152_02405, MED152_02435, MED152_02810, MED152_03055, MED152_03355, MED152_07660, MED152_08085, MED152_10165, MED152_13039
Cold-shock DNA-binding protein (3)	Cryoprotection	MED152_01260, MED152_02750, MED152_06655
Stationary-phase survival acid phosphatase SurE	Starvation	MED152_06290
Guanosine polyphosphate (ppGpp) pyrophosphohydrolase/synthetase	Starvation	MED152_01945
Phosphate starvation-inducible protein, PhoH	Starvation	MED152_13139
Catalase	Oxidative	MED152_05245
Cu/Zn superoxide dismutase	Oxidative	MED152_04650
Fe/Mn superoxide dismutase	Oxidative	MED152_07780
Cytochrome-c peroxidase (2)	Oxidative	MED152_03805, MED152_08020
Protein-Met-S-oxide reductase, type A (3)	Oxidative	MED152_03040, MED152_07435, MED152_07450
Protein-Met-S-oxide reductase, type B (2)	Oxidative	MED152_01175, MED152_07430
Glutathione peroxidase	Oxidative	MED152_05385
AhpC/TSA family protein (3)	Oxidative	MED152_02100, MED152_05240, MED152_09055
Redoxin (16)	Oxidative	MED152_00520, MED152_00525, MED152_01395, MED152_01805, MED152_01910, MED152_03415, MED152_05235, MED152_06110, MED152_06230, MED152_06900, MED152_07620, MED152_08425, MED152_10120, MED152_10405, MED152_10550, MED152_11584
Arsenate oxidoreductase ArsC (3)	Metals	MED152_00320, MED152_02710, MED152_07715
Arsenite resistance protein ArsB	Metals	MED152_02140
Heavy metal transport/detoxification protein (2)	Metals	MED152_09525, MED152_11224
MerTP Mercuric Ion (Hg^{2+}) Permease (MerTP) family transporter	Metals	MED152_00110*
Arsenical Resistance-3 (ACR3) family transporter	Metals	MED152_02140*
Mn^{2+} Transporter (Nramp) family (2)	Metals	MED152_03840*, MED152_11859*
Integral membrane protein TerC family (2)	Metals	MED152_01215, MED152_13169
β-Lactamase (8)	Drugs	MED152_03350, MED152_03530, MED152_07260, MED152_07965, MED152_08005, MED152_09070, MED152_10295, MED152_10385
Chloramphenicol acetyltransferase	Drugs	MED152_03570
ATP-Binding Cassette (ABC) type transporter (4)	Drugs	MED152_00590*, MED152_00620*, MED152_11514*, MED152_13184*
Major Facilitator Superfamily (MFS) transporter (2)	Drugs	MED152_08705*, MED152_11639*
Multidrug/Oligosaccharidyl-lipid/Polysaccharide (MOP) Flippase superfamily transporter (3)	Drugs	MED152_01915*, MED152_02820*, MED152_09425*
Resistance-Nodulation-Cell Division (RND) superfamily transporter	Drugs	MED152_08525*

ORFs with an asterisk also appear in Table S8 showing transporters.

Table S4. Genes coding for replication, repair, and recombination mechanisms in MED152

Gene name	ORFs	Protein description; pathway
uvrA	MED152_01780, MED152_04900	ATPase, DNA binding; nucleotide excision repair
uvrB	MED152_03455	Helicase; nucleotide excision repair
uvrC	MED152_09500	Nuclease; nucleotide excision repair
uvrD	MED152_01615, MED152_10775	UvrD/REP helicase; nucleotide excision repair
mfd	MED152_10740	Transcription-repair coupling factor; nucleotide excision repair
ung	MED152_00800	Uracyl DNA glycosylase; base-excision repair
nth	MED152_01800, MED152_09640	HhH-GPD superfamily; base excision repair
tag	MED152_11989	DNA-3-methyladenine glycosidase I; base excision repair
xth	MED152_10450	Exodeoxyribonuclease III; base excision repair
	MED152_00700	Non-canonical purine NTP pyrophosphatase, rdgB/HAM1 family; DNA repair
radA	MED152_08245	DNA repair protein RadA/Sms
radC	MED152_12379	DNA repair protein RadC
mutL	MED152_01325	DNA mismatch repair protein
mutS	MED152_00795, MED152_08905, MED152_09620	DNA mismatch repair protein
recA	MED152_08735	Recombinase; DNA recombination
recD	MED152_00290	ATP-dependent exoDNAse (exonuclease V), alpha subunit; DNA recombination
recF	MED152_01345	DNA replication; DNA recombination
recJ	MED152_07910	Single-stranded-DNA-specific exonuclease; DNA degradation
recN	MED152_13259	DNA repair protein RecN
recO	MED152_10500	Recombination protein RecO
recR	MED152_04560	Recombination protein RecR
ruvA	MED152_06690	DNA helicase RuvA; DNA recombination
ruvB	MED152_10285, MED152_12424	DNA helicase RuvB; DNA recombination
ruvC	MED152_12024	Endodeoxyribonuclease RuvC; DNA recombination
	MED152_00210, MED152_01115, MED152_02515, MED152_04860, MED152_07400, MED152_09890	DNA recombination
cinA	MED152_03985	Competence/DNA damage-inducible protein CinA
	MED152_00230, MED152_02730, MED152_05735, MED152_07035, MED152_09645	DNA binding protein
polA	MED152_10105	DNA polymerase I
holA	MED152_00675	DNA polymerase III, delta subunit
dnaA	MED152_13404	Chromosomal replication initiator protein DnaA
dnaB	MED152_01985	DnaB replicative DNA helicase
dnaE	MED152_07245, MED152_11814	DNA polymerase III, alpha subunit
dnaG	MED152_11074	DNA primase
dnaN	MED152_04670	DNA polymerase III, beta subunit
dnaQ	MED152_01490, MED152_05060, MED152_11754	DNA polymerase III, epsilon subunit
dnaX	MED152_03045	DNA polymerase III, subunit gamma/tau
holB	MED152_09190	DNA polymerase III, delta' subunit
ligA	MED152_13454	DNA ligase, NAD-dependent
topA	MED152_11284	DNA topoisomerase I
exc	MED152_01555	DNA topoisomerase III
parC	MED152_01590	DNA topoisomerase IV, subunit A
parE	MED152_01585	DNA topoisomerase IV, subunit B
gyrA	MED152_04790	DNA gyrase, subunit A
gyrB	MED152_12139	DNA gyrase, subunit B
ssb	MED152_09160, MED152_09635	Single-stranded DNA-binding protein
rnhB	MED152_05995	Ribonuclease HII
	MED152_09655	Ribonuclease, Rne/Rng family
mfd	MED152_10740	Transcription-repair coupling factor
dcm	MED152_07385	DNA (cytosine-5-)-methyltransferase; restriction/modification
ogt	MED152_07670	Methylated DNA-protein Cys methyltransferase
	MED152_09340	6-O-methylguanine DNA methyltransferase
	MED152_07920	Helicase
cspD	MED152_01260, MED152_02750, MED152_06655	'Cold-shock' DNA-binding protein
	MED152_01310, MED152_04910, MED152_07835, MED152_09660, MED152_11354	Endonuclease/exonuclease/phosphatase family protein
	MED152_00240, MED152_01675, MED152_01850, MED152_02405, MED152_02435, MED152_02810, MED152_03055, MED152_03355, MED152_07660, MED152_08085, MED152_10165, MED152_13039	DEAD/DEAH box helicase
	MED152_05340, MED152_05350, MED152_05405	DNA photolyase/cryptochrome
	MED152_12919, MED152_13194	NUDIX hydrolase; base-excision repair

Table S5. Distribution of the anaplerotic enzymes pyruvate and PEP carboxylases, as well as carbonate anhydrase in cultured marine heterotrophic bacteria that contain the proteorhodopsin gene

Organism	Genbank accession no.	Genome size, bp	Taxonomy	Pyruvate carboxylase	PEP carboxylase	Carbonate anhydrase
Polaribacter sp. MED152	AANA00000000	2,967,150	Bacteroidetes	+	+	+
Polaribacter irgensii 23-P	AAOG00000000	2,745,458	Bacteroidetes	+	+	+
Dokdonia sp. MED134	AAMZ00000000	3,301,953	Bacteroidetes	+	+	+
Flavobacterium BAL38	AAXX00000000	2,806,989	Bacteroidetes	-	+	+
Pelagibacter ubique HTCC1002	AAPV00000000	1,328,618	Alphaprot.	-	-	-
Pelagibacter ubique HTCC1062	CP000084	1,308,759	Alphaprot.	-	+	-
Octadecabacter antarcticus 307	Moore F.	4,909,025	Alphaprot.	+	-	+
BAL199	ABHC00000000	6,162,950	Alphaprot.	-	+	++
Methylophilales HTCC2181	AAUX00000000	1,304,428	Betaprot.	-	-	+
Vibrio harveyi ATCC BAA-111	CP000789	6,058,377	Gammaprot.	-	+	+
Vibrio sp. S14	AAOJ00000000	5,169,214	Gammaprot.	-	+	+
HTCC2143	AAVT00000000	3,925,629	Gammaprot.	-	+	++
HTCC2207	AAPI00000000	2,620,870	Gammaprot.	-	-	+
Photobacterium sp. SKA34	AAOU00000000	4,991,572	Gammaprot.	-	+	+
Marinobacter sp. ELB17	AAXY00000000	4,928,595	Gammaprot.	-	+	++

The presence of the proteorhodopsin gene is based on hits against PF01036. Further, only those peptides that contained an Asp85 and Asp96, or a different carboxylate residue, were considered (Asp85 and Asp96 functions as the H^+ acceptor and donor, respectively, during the rhodopsin photocycle in the H^+ pump type of rhodopsins). Number of pluses indicates the number of genes with the same putative function. Pyruvate carboxylase (EC 6.4.1.1) catalyzes the following reaction: pyruvate + ATP + HCO_3^- <=> ADP + phosphate + oxaloacetate. PEP carboxylase (EC 4.1.1.31) catalyzes the reaction: PEP + HCO_3^- + H_2O <=> phosphate + oxaloacetate. Carbonate anhydrase (EC 4.2.1.1) catalyzes the reaction: HCO_3^- <=> CO_2 + H_2O.

Table S6. Domains and peptides with a putative role in light absorption and response

Domain/peptide	ORFs	Function
Light sensing		
Phytochrome region	MED152_00100	Red and far-red light sensor; chromophore is a linear tetrapyrrole bound via a Cys residue
PAS and PAC domains	MED152_00100, MED152_03150, MED152_04250, MED152_06325	Flavin-binding; component of phytochromes; sensor of oxygen, redox and light
GAF domain	MED152_00100, MED152_02465, MED152_03150, MED152_05910, MED152_11204	Binds small-molecules (cAMP or cGMP) that act as second messengers; component of phytochromes
DNA photolyase/cryptochrome, animal cryptochrome and (6–4) photolyase family	MED152_05350	Flavin-binding; senses short wavelengths of visible light
Cryptochrome, DASH family	MED152_05340	Flavin-binding; senses short wavelengths of visible light
Deoxyribodipyrimidine photo-lyase, class I	MED152_05405	Flavin-binding; senses short wavelengths of visible light
BLUF domain	MED152_08960	Flavin-binding; blue light sensor
Synthesis of β-carotene		
Isopentenyl-diphosphate delta-isomerase (idi)	MED152_03605	Synthesis of β-carotene from isopentenyl diphosphate and dimethylallyl diphosphate
Farnesyl-diphosphate synthase (ispA)	MED152_07135 or MED152_08510	
Geranylgeranyl diphosphate synthase (crtE)	MED152_07135 or MED152_08510	
Phytoene synthase (crtB)	MED152_02670	
Phytoene desaturase (crtI)	MED152_02575 or MED152_02675	
Lycopene β-cyclase (crtY)	MED152_02660	
15,15'-β-Carotene dioxygenase (blh)	MED152_12849	Final step in the synthesis of retinal from β-carotene
Synthesis of proteorhodopsin apoprotein		
Opsin	MED152_12844	
Additional carotenoids		
Spheroidene monooxygenase (crtA)	MED152_02580	
Methoxyneurosporene dehydrogenase (crtD)	MED152_02575 or MED152_02675	
β-Carotene hydroxylase (crtZ)	MED152_02565 or MED152_02665	Synthesis of zeaxanthin from ß-carotene

Table S7. Distribution of putative light sensors in proteorhodopsin-containing marine bacteria (as defined in Table S5)

Organism	BLUF	Phytochrome	Animal cryptochrome and (6–4) photolyase family	Cryptochrome, DASH family	Photolyase, class I
Polaribacter sp. MED152	+	+	+	+	+
Polaribacter irgensii 23-P	-	-	++	+	+
Dokdonia sp. MED134	+	+	+	+	+
Flavobacterium BAL38	-	-	+	+	+
P. ubique HTCC1002	-	-	-	-	+
P. ubique HTCC1062	-	-	-	-	+
Octadecabacter antarcticus 307	-	-	+	-	+
Methylophilales HTCC2181	-	-	-	+	+
Vibrio harveyi ATCC BAA-1116	-	-	+	+	+
Vibrio sp. S14	-	-	+	+	+
HTCC2143	-	-	+	+	+
HTCC2207	-	-	-	+	+
Photobacterium sp. SKA34	-	-	+	+	+
Marinobacter sp. ELB17	-	-	++	+	+

Presence of light sensors are predicted by hits to specific PFAMs or TIGRFAMs. DNA photolyase/cryptochrome subfamilies are based on phylogenetic analysis as in Fig. S6, as well as on custom-built PFAMs. Of 80 Moore genomes, 30 contain at least one BLUF domain, 18 contain at least one phytochrome domain, and 12 contain three cryptochrome/photolyase peptides (most contain one or two and no other contain four cryptochrome/photolyase peptides).

Table S8. Transporters identified in the genome of MED152

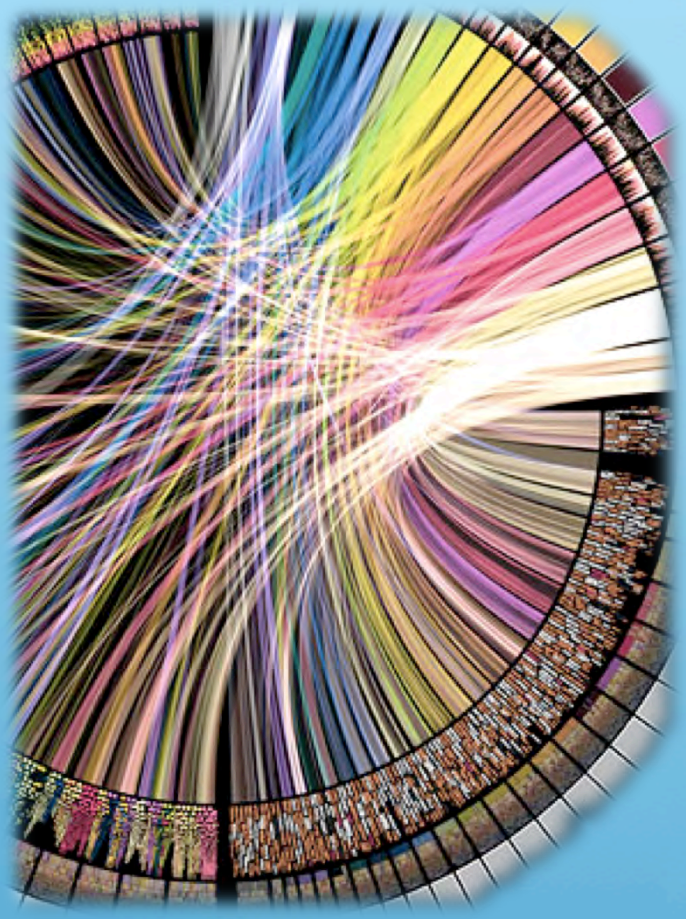
Substrate	Type of transporter	ORFs
	Ion channels	
Cl^-	Chloride Channel (ClC) family	MED152_07855
Mg^{2+}/Co^{2+}	CorA Metal Ion Transporter (MIT) family	MED152_12054
	Large Conductance Mechanosensitive Ion Channel (MscL) family	MED152_07755, MED152_08290
	Small Conductance Mechanosensitive Ion Channel (MscS) family	MED152_03080, MED152_05265, MED152_06730, MED152_10700, MED152_11349, MED152_12279
K^+	Voltage-Gated Ion Channel (VIC) superfamily	MED152_01430, MED152_04890
	ATP-Binding Cassette (ABC) superfamily	
bacitracin		MED152_00590
multidrug		MED152_00620
Fe^{3+}		MED152_00695, MED152_01595, MED152_11574
unknown		MED152_01685
unknown		MED152_02455
unknown		MED152_02840, MED152_02845, MED152_02850
lipoprotein releasing		MED152_03470, MED152_03875
lipid A export		MED152_04055
lipoprotein releasing		MED152_04345
unknown		MED152_04415
Na^+ efflux		MED152_05270, MED152_05275
Fe-S assembly/SufBCD system		MED152_07995, MED152_08000, MED152_08010
cell division		MED152_12239
lipoprotein releasing		MED152_10825
multidrug		MED152_11514
toluene tolerance		MED152_03275, MED152_12664, MED152_12669
gliding motility		MED152_11809
gliding motility		MED152_13119
multidrug		MED152_13184
Na^+ efflux		MED152_13214
	Membrane pyrophosphatase	
H^+		MED152_11924
	P-type ATPase superfamily	
Cu^{2+}		MED152_00135
Zn^{2+}		MED152_00155
Cation		MED152_03205
	Secondary transporters	
$H_2AsO_3^-$	Arsenical Resistance-3 (ACR3) family	MED152_02140
Na^+/alanine	Ala or Gly:Cation Symporter (AGCS) family	MED152_04880
NH_4^+	Ammonium Transporter (Amt) family	MED152_05800, MED152_05810
amino acid	Amino Acid-Polyamine-Organocation (APC) family	MED152_09705
Na^+/unknown	Bile Acid:Na^+ Symporter (BASS) family	MED152_05040
Na^+/Ca^{2+}	Ca^{2+}:Cation Antiporter (CaCA) family	MED152_11634
nucleosides	Concentrative Nucleoside Transporter (CNT) family	MED152_11959
Na^+/H^+	Monovalent Cation:Proton Antiporter-1 (CPA1) family	MED152_02960, MED152_03050
K^+	Monovalent Cation:Proton Antiporter-2 (CPA2) family	MED152_02125
H^+/Na^+:glutamate	Dicarboxylate/Amino Acid:Cation (Na^+ or H^+) Symporter (DAACS) family	MED152_02010
Na^+/dicarboxylate	Divalent Anion:Na^+ Symporter (DASS) family	MED152_09770
drug/metabolite	Drug/Metabolite Transporter (DMT) Superfamily	MED152_02080, MED152_09260, MED152_09315, MED152_09530, MED152_10735
H^+/Na^+:sugar	Glycoside-Pentoside-Hexuronide (GPH):Cation Symporter family	MED152_00365, MED152_00460, MED152_05120
branched-chain amino acid	Branched Chain Amino Acid:Cation Symporter (LIVCS) family	MED152_08060
unknown	Major Facilitator Superfamily (MFS)	MED152_01190, MED152_03745, MED152_05095
nucleosides	Major Facilitator Superfamily (MFS)	MED152_05175
glucose/galactose	Major Facilitator Superfamily (MFS)	MED152_08460, MED152_13289
multidrug efflux	Major Facilitator Superfamily (MFS)	MED152_08705, MED152_11639
arabinose	Major Facilitator Superfamily (MFS)	MED152_09090

Table S8 ***cont.***

Substrate	Type of transporter	ORFs
hexuranate	Major Facilitator Superfamily (MFS)	MED152_09855
multidrug efflux	Multidrug/Oligosaccharidyl-lipid/ Polysaccharide (MOP) Flippase superfamily	MED152_01915, MED152_02820, MED152_09425
oligosaccharide repeat unit	Multidrug/Oligosaccharidyl-lipid/ Polysaccharide (MOP) Flippase superfamily	MED152_03130, MED152_10600
polysaccharide	Multidrug/Oligosaccharidyl-lipid/ Polysaccharide (MOP) Flippase superfamily	MED152_04985
O-antigen	Multidrug/Oligosaccharidyl-lipid/ Polysaccharide (MOP) Flippase superfamily	MED152_12399
Na^+/H^+	NhaC $Na^+:H^+$ Antiporter (NhaC) family	MED152_06020, MED152_08675
Na^+/H^+	NhaD $Na^+:H^+$ Antiporter (NhaD) family	MED152_05770
Mn^{2+}	Metal Ion (Mn^{2+}-iron) Transporter (Nramp) family	MED152_03840, MED152_11859
60-kDa inner membrane protein OxaA homolog	Cytochrome Oxidase Biogenesis (Oxa1) family	MED152_13299
dipeptide/tripeptide: H^+	H^+-dependent Oligopeptide Transporter (POT) family	MED152_06255, MED152_06260
amino acid	Resistance to Homoserine/Thr (RhtB) family	MED152_01410
Cu^{2+}/Ag^+ efflux	Resistance-Nodulation-Cell Division (RND) superfamily	MED152_00145
multidrug efflux	Resistance-Nodulation-Cell Division (RND) superfamily	MED152_08525
cation/multidrug efflux	Resistance-Nodulation-Cell Division (RND) superfamily	MED152_10530
unknown	Resistance-Nodulation-Cell Division (RND) superfamily	MED152_11044
protein-export (SecDF)	Resistance-Nodulation-Cell Division (RND) superfamily	MED152_12119
Na^+/HCO_3^-	Na^+-dependent Bicarbonate Transporter (SBT) family	MED152_03855
Na^+/solute	Solute:Na^+ Symporter (SSS) family	MED152_04565, MED152_04660, MED152_05085
$SO_4^=$	$SO_4^=$ Permease (SulP) family	MED152_09030, MED152_10290, MED152_12914
protein export	Twin Arginine Targeting (Tat) family	MED152_02445, MED152_08785
K^+	K^+ Transporter (Trk) family	MED152_12369
	Unclassified transporters	
Ca^{2+}	Ca^{2+}-ATPase	MED152_12659
Fe^{2+}	Fe^{2+} Uptake (FeoB) family	MED152_06540
Hg^{2+}	MerTP Mercuric Ion (Hg^{2+}) Permease (MerTP) family	MED152_00110
Mg^{2+}	Mg^{2+} Transporter-E (MgtE) family	MED152_08915
nicotinamide mononucleotide	Nicotinamide Mononucleotide (NMN) Uptake Permease (PnuC) family	MED152_09745

Table S9. Genes and domains with a potential role in adhesion

Domain	ORFs	Protein family
Fasciclin	MED152_12639, MED152_13164	PF02469
Thrombospondin type 3 repeat	MED152_00310, MED152_03360, MED152_04530, MED152_10785, MED152_07970	PF02412
Fibronectin type III	MED152_07970	PF00041
PKD	MED152_00430, MED152_03360, MED152_04535, MED152_04570, MED152_04575, MED152_09845	PF00801
FG-GAP	MED152_05130, MED152_05135, MED152_05145, MED152_08145	PF01839
ASPIC and UnbV	MED152_05145, MED152_08145, MED152_13274	PF07593
Lectin C-type domain	MED152_07900	PF00059
Cadherin	MED152_00085	IPR002126
HYR	MED152_00085	PF02494
Leu rich repeat	MED152_10960, MED152_11029	PF00560
β-Propellor repeat	MED152_02285	IPR013519
Cellulose binding motif	MED152_00085	PF00553
DUF11	MED152_07970	PF01345
Von Willebrand factor type A	MED152_08995, MED152_09005, MED152_09010	PF00092
Cna protein B-type	MED152_10890	PF05738
Carbohydrate binding domain	MED152_00430, MED152_03360, MED152_04575, MED152_09845, MED152_11564	PF02018
Carbohydrate binding module (family 6)	MED152_00755	PF03422
Cellulose binding domain	MED152_00870	PF00942
DUF638	MED152_03735	PF04829
F5/8 type C domain	MED152_08630, MED152_13279	PF00754
Starch-binding domain	MED152_11969	PF00686



Chapter 2

Comparative genomics of marine Bacteroidetes with and without proteorhodopsin

Beatriz Fernández-Gómez, Michael Richter, Margarete Schüler, Jarone Pinhassi, Silvia G. Acinas, José M. González, and Carlos Pedrós-Alió

ABSTRACT

Analysis of marine Bacteroidetes genomes reveals a life strategy different from those of other important phyla of marine bacterioplankton such as Cyanobacteria and Proteobacteria. Comparative genome analysis of four cultured representatives of the Bacteroidetes phylum indicates that these bacteria have many adaptations to grow attached to particles and degrading polymers, including a large number of peptidases, glycoside hydrolases, glycosyl transferases, adhesion proteins, and the genes for gliding motility. Several of the genes are located in close association with genes for TonB-dependent-receptors and transducers, suggesting an integrated regulation of adhesion and degradation of polymers. Genes shared among the four Bacteroidetes genomes include those for signal transduction, transport, motility and a large percentage of membrane and secreted proteins. In addition, the proteorhodopsin-containing marine Bacteroidetes seem to change their metabolism in order to survive when floating freely in the nutrient-poor surface ocean. The genes potentially important in this alternative life style include many genes involved in the anaplerotic fixation of CO_2. The analysis also shows that marine Bacteroidetes are specialized in the use of polymeric organic matter and, particularly, proteins.

INTRODUCTION

Members of the phylum Bacteroidetes constitude the most abundant group of bacteria in the ocean after Proteobacteria and Cyanobacteria (Glöckner *et al.,* 1999; Kirchman 2002; Amaral-Zettler *et al.,* 2010). They account for a significant fraction of marine bacterioplankton especially in coastal areas, where they represent between 10% and 30% of the total bacterial counts (Alonso-Sáez and Gasol 2007). They are globally distributed (Pommier *et al.,* 2007) and it is possible to find them in a variety of marine environments such as coastal, offshore, sediments, and hydrothermal vents (Alonso *et al.,* 2007; Pommier *et al.,* 2007). In a recent analysis of metagenomes from the Global Ocean Sampling, Yooseph *et al.* (2010) analyzed which genomes from cultured microorganisms recruited the most fragments from the metagenomes. *Polaribacter* sp. MED152, one of the Bacteroidetes analyzed here, was the 10th genome that recruited the most fragments. Only the genomes of four *Prochlorococcus*, three *Pelagibacter,* one *Synechoccocus*, and one Rhodobacterales bacterium were covered at greater depth than *Polaribacter* sp. MED152. Finally, a CARD-FISH study of Bacteroidetes across the orth Atlantic Ocean showed *Polaribacter* to be the most abundant Bateroidetes in most samples (Gómez-Pereira *et al*., 2010). This shows that the genus is widely distributed and abundant in the oceans.

The better known members of the Bacteroidetes are specialized in processing polymeric organic matter, particularly in the mammalian gut (for example *Bacteroides* spp.) or in soils (*Cytophaga*). In aquatic habitats they are abundant during and following algal blooms (Pinhassi *et al.,* 2004), showing a preference for consuming polymers rather than monomers (Cottrell and Kirchman 2000). Thus, they likely present a different life strategy to that of other marine bacteria such as the Alphaproteobacteria and the Cyanobacteria and therefore, they will probably fill different niches. Cyanobacteria are photoautotrophs. Alphaproteobacteria are aerobic heterotrophs that preferentially use monomers and live suspended in the water column. If the preference of Bacteroidetes for polymers and an existence attached to surfaces could be confirmed, their role in the carbon cycle of the oceans would be complementary to that of the other two groups.

Analysis of the first marine Bacteroidetes genome sequence, that of *Gramella forsetii* KT0803 (Bauer *et al.,* 2006), revealed a 3.8 Mb genome with a relatively large number of glycoside hydrolases, peptidases, and adhesion proteins, suggesting that *G. forsetii* was well adapted to degrade high molecular weigh compounds. These traits were also observed in *Polaribacter* sp. MED152, the second free-living marine Bacteroidetes genome analyzed (Gónzalez *et al.,* 2008). The latter authors proposed a dual life strategy for *Polaribacter*. This

organism would grow optimally attached to particles or other surfaces where it could move by gliding motility searching for polymeric substances, and degrade them using its large array of peptidases and glycoside hydrolases. When labile organic matter were exhausted on a particle, *Polaribacter* would be forced to float passively in the nutrient poor water column in search of another particle. Under these conditions, it would use light to obtain energy and, thus, optimize the use of whatever little organic matter it could find. The genome of another marine flavobacterium, *Dokdonia* sp. MED134 (González *et al.*, 2011), suggests that this PR-containing bacterium has similar characteristics to those of *Polaribacter*.

Several more genomes have been sequenced since these studies were published and it should now be possible to test whether the assumed role of Bacteroidetes as degraders of polymers can be confirmed. We also wanted to check wether the strategy proposed for *Polaribacter* holds with other PR-containing bacteria. In this paper we present a comparative genome analysis using the available genomes of marine Bacteroidetes but focusing on four cultured coastal Bacteroidetes strains, *Gramella forsetii* KT0803 from the North Sea, and *Polaribacter* sp. MED152, *Dokdonia* sp. MED134 and *Leeuwenhoekiella blandensis* MED217 from the Mediterranean Sea. *Polaribacter* and *Dokdonia* have a gene coding for proteorhodopsin (PR) while the other two do not. We analyzed which genetic features identified in *Polaribacter* are restricted to PR-containing bacteria (around 15 genomes), which ones are common to marine Bacteroidetes (21 genomes), and which ones are present in all Bacteroidetes (28 genomes).

MATERIAL AND METHODS

Isolation of Bacteroidetes

Polaribacter sp. MED152, *Dokdonia* sp. MED134 and *L. blandensis* MED217 were isolated in 2001 from Northwest Mediterranean Sea surface water (0.5 m depth) collected 1 km off the coast at the Blanes Bay Microbial Observatory (BBMO, http://www.icm.csic.es/bio/projects/icmicrobis/bbmo/), Spain (41º 40' N, 2º 48' E). All three were isolated on ZoBell agar plates. *Gramella forsetii* KT0803 was isolated in 1999 in the North Sea surface water (1 m depth) collected 1 km off the coast of the island of Helgoland, Germany (54º 09' N, 7º 52' E). It was isolated on MPM medium plates (Schut *et al.,* 1993).

Genome sequencing and assembly

Whole-genome sequencing of *Polaribacter* sp. MED152, *Dokdonia* sp. MED134 and *L. blandensis* MED217 was carried out by the J. Craig Venter Institute through the Gordon and Betty Moore Foundation initiative in Marine Microbiology. A Sanger/pyrosequencing hybrid method was used (Goldberg *et al.,* 2006). The complete genome sequence of *Polaribacter sp*. MED152 contained 30,151 reads, with an average of 9.6-fold sequence coverage and has been described in González *et al.* (2008). The complete genome sequence of *Dokdonia sp.* MED134 contained 33,882 reads, with an average of 9.8-fold sequence coverage (González *et al.,* 2011). The complete genome sequence of *L. blandensis* MED217 contained 47,494 reads, with an average of 10-fold sequence coverage. Sequencing of *Gramella forsetii* KT0803 was carried out at the Max Planck Institute for Molecular Genetics (Berlin). It consists of 64,653 high quality reads with 9.8-fold sequence coverage and has been described in (Bauer *et al.,* 2006).

Gene prediction and annotation

Data mining was carried out with GenDB v2.2 (Meyer *et al.,* 2003), supplemented by the tool JCoast (Richter *et al.,* 2008). For each predicted ORF observations were collected from similarity searches against sequence databases NCBI-nr, Swiss-Prot, KEGG and genomes DB (Richter *et al.,* 2008), and for protein databases Pfam (Finn *et al.,* 2008) and InterPro (Hunter *et al.,* 2009). SignalP was used for signal peptide predictions (Bendtsen *et al.,* 2004) and TMHMM for transmembrane helix-analysis (Krogh *et al.,* 2001). The annotation of *Polaribacter* sp. MED152 (González *et al.,* 2008), *Gramella forsetii* KT0803 (Bauer *et al.,* 2006), and *Dokdonia sp*. MED134 (González *et al.,* 2011) genomes was manually curated and refined for

each ORF, while the *L. blandensis* MED217 genome remains automatically annotated.

Comparative analysis

Comparative analysis and data mining were carried out using JCoast (Richter *et al.,* 2008). Paralogous protein families within each of the four genomes were identified by BLASTP (Altschul *et al.,* 1990) all-against-all similarity comparisons at a significance cut-off level of 10^{-10}. Extracted proteins were grouped into families and functional proteins were classified according to Pfam and Cluster of Orthologous Genes, COG (Tatusov *et al.,* 1997).

Comparison of the shared gene content by reciprocal best matches (RBMs)

Determination of the shared gene content was done by a pair-wise BLASTp all versus all search between investigated organisms. Reciprocal best matches were counted by a BLASTP result with an expectation values of $E<10^{-5}$ each, and a subject coverage of over 65%.

Searching for genes and domains in the genomes

Signal transduction proteins, transporters, domains related to surface adhesion and other genes needed for comparison were found using the Pfam search included in JCoast (Richter *et al.,* 2008), with a threshold of $E\leq10^{-4}$.

Phylogenetic analysis

Maximum likelihood trees were constructed with full-length 16S rRNA sequences from forty-seven completed genomes of Bacteroidetes. After the alignment of these sequences using SINA Aligner offered by the Silva database (Pruesse *et al.,* 2007), the poorly aligned positions and divergent regions of the alignment were eliminated with the computer program Gblocks (Castresana 2000). The trees were constructed using RAxML server (Stamatakis *et al.,* 2008) with the GTR nucleotide substitution model. The online tool Interactive Tree of Life (iTOL) (Letunic and Bork 2007, 2011) was used for editing and presentation of the tree.

Accession Numbers

The complete *Gramella forsetii* KT0803 sequence is available under GenBank accession number CU207366. *Polaribacter* sp. MED152, *Dokdonia* sp. MED134 and *L. blanden-*

sis MED217 sequences are available under GenBank accession numbers AANA00000000, AAMZ00000000 and AANC00000000, respectively.

RESULTS AND DISCUSSION

1. PR and genome size

The basic properties of the four genomes are shown in Table 1. Several tables in supplementary information summarize a comparison of the genes or domains identified in the four genomes concerning nitrogen, phosphorous, or sulfur acquisition (Suppl. Tables 1, 2 and 3), sodium transporters (Suppl. Table 4), one- and two-component systems (Suppl. Tables 5 and 6), adhesion (Suppl. Table 7), paralogous families (Suppl. Table 8), and clusters of polymer degradation genes (Suppl. Table 9).

Table 1. General features of the four Bacteroidetes genomes analyzed.

Characteristic	*Polaribacter* MED152[a]	*Dokdonia* MED134[b]	*Gramella* KT0803[c]	*Leeuwenhoek.* MED217[d]
Size (Mega base pairs)	2 967	3 302	3 799	4 244
G+C content (%)	30.7	38.2	36.6	39.8
Protein-coding genes	2 646	3 008	3 585	3 689
Coding density (%)	92.9	89.9	90.2	90.6
Stable RNAs:				
rRNA operons	3	3	3	4
tRNA	37	43	44	38
Paralogous Families	246	245	356	351
Genes in P. Families (%)	27.2	25.2	33.3	32.8
Single Copy Genes (%)	72.8	74.8	66.6	67.2
Phage-integrase like genes	3	1	2	18
Transposases	0	2	68	24
Conjugative transposon	No	Yes	Yes	Yes

[a] Manually annotated genome published in González *et al*. (2008)
[b] Manually annotated genome published in González *et al.* (2011)
[c] Manually annotated genome published in Bauer *et al*. (2006)
[d] Bacterium described in Pinhassi *et al*. (2006). The automatically annotated genome is available in GenBank.

The four genomes ranged in size between 2.97 and 4.24 Mb (Table 1). The two bacteria without PR (PR- from now on) had larger genomes (*G. forsetii* and *L. blandensis*) than the two with PR (PR+ from now on).

Table 2. Genes shared between each pair of the four Bacteroidetes genomes analyzed based on reciprocal best matches.

	MED217	KT0803	MED134	MED152
MED217		2122	1985	1741
KT0803			1888	1692
MED134				1762

The four genomes were compared pairwise by reciprocal best matches. The two larger genomes shared 2122 orthologous genes while the two smaller genomes shared 1762 (Table 2). In all cases, shared genes accounted for between 50 and 59% of the genome.

There were at least three reasons why two of the genomes were larger. First, they had more paralogous families and a larger percent of genes in such families (Table 1 and Fig. 1). Gene duplication and subsequent modifications to carry out novel functions by paralogous proteins is a well-known mechanism of bacterial evolution (see a recent review in Andersson *et al*., 2009).

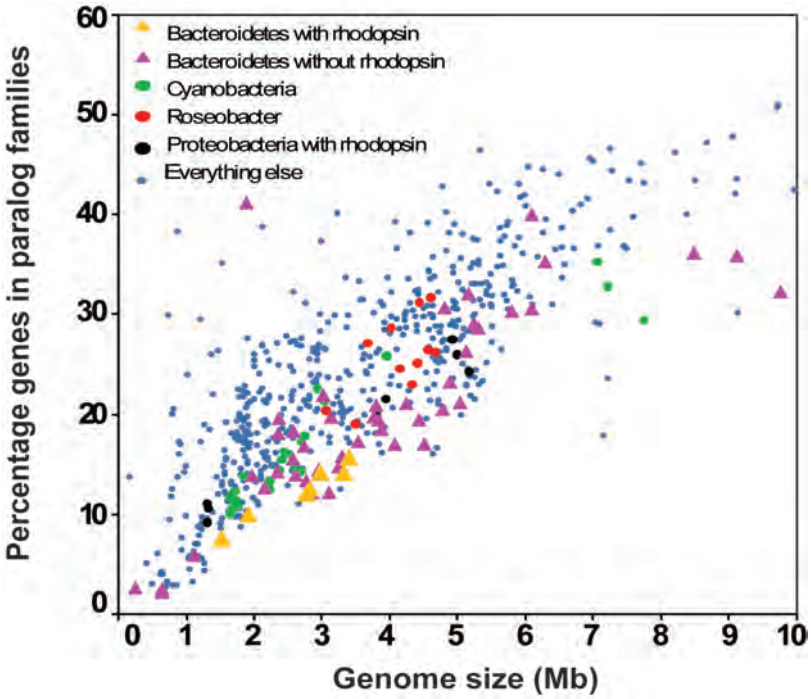


Figure 1. Percentage of genes in paralogous families versus genome size for a number of bacteria. The four PR-Bacteroidetes genomes smaller than 2 Mb correspond to endosymbionts of arthropods.

Second, the larger genomes also had more transposases and at least one conjugative transposon (Table 1). In effect, we tested the presence of genomic islands (GIs) in a large set of marine bacteria (Fernández-Gómez *et al.,* submitted) with the software IslandViewer (Languille and Brinkman 2009, Languille *et al*., 2010), and the MED217 genome had eight, three of them larger than 9.5 kb with a total of 70.5 kb, representing 1.18% of the genome. *Gramella forsetii* had five GIs with a total of 105 kb)MED134 and MED152, on the other hand, only had two and one GIs (with a total of 39 kb and 12 kb respectively). Some of the genes present in MED217 and missing in the two smaller genomes were actually located within genomic islands. Thus, missing areas labeled A to H in Figure 2A corresponded to genomic islands detected in MED217 (Fig. 2B). Genomic island H, for example, had integrases and transposases at both ends of the island, and CRISPRs at one end. The latter are repetitive palindromic sequences that, in association with genes *cas*7, *cas*5 and two *cas*3 (also present), serve a defensive function against phages and can be mobilized by horizontal gene transfer (HGT) (Haft *et al.,* 2005). Most of the annotated genes within the island coded for hypothetical proteins (up to 55% of the genes within GIs) as previously reported for other marine bacteria (Hsiao *et al.,* 2005). However, a regulatory sigma factor and a putative nitrite reductase were also found in the H genomic island of MED217. Interestingly, a nitrite reductase has been found in the deep-sea marine flavobacterium *Zunongwangia profunda* SM-A87 (Qin *et al.,* 2010). A comparison between the nitrite reductases of MED217 and *Zunongwangia* showed a low percentage of amino acid identity (24%). Actually, the best hits of the MED217 nitrite reductase sequence were against gammaproteobacterial proteins genes, suggesting a HGT origin of this function in MED217 (data not shown). The coincidence of GIs with genes absent from the smaller genomes indicates that one of the mechanisms responsible for the differences in size among the genomes was HGT.

Third, some genes were present in the larger genomes that code additional functions absent from the smaller ones. The area labeled Z in Fig. 2, for example, was present in the two large genomes and absent from both small genomes, but it did not coincide with any GI. The genes in this area were mostly involved in the metabolism of sugars, particularly arabinose (Fig. 3): they included the three structural genes of the arabinose operon (*ara*A, *ara*D, and *ara*B, that are responsible for converting arabinose into D-xylulose-5P), *gal*M (codes an epimerase capable of interconverting L- and D- arabinose as well as other sugars), and a few genes also related to sugar metabolism, including a sodium/glucose co-transporter. Presence of these genes should allow the bacterium to use arabinose by converting it to D-xylulose-5P, which is then transformed into D-ribulose-5P by a widely distributed enzyme (pentose-5-phosphate 3-epimerase) and can then enter the pentose phosphate pathway (present in the four genomes). In effect, when utilization of arabinose was experimentally tested with the

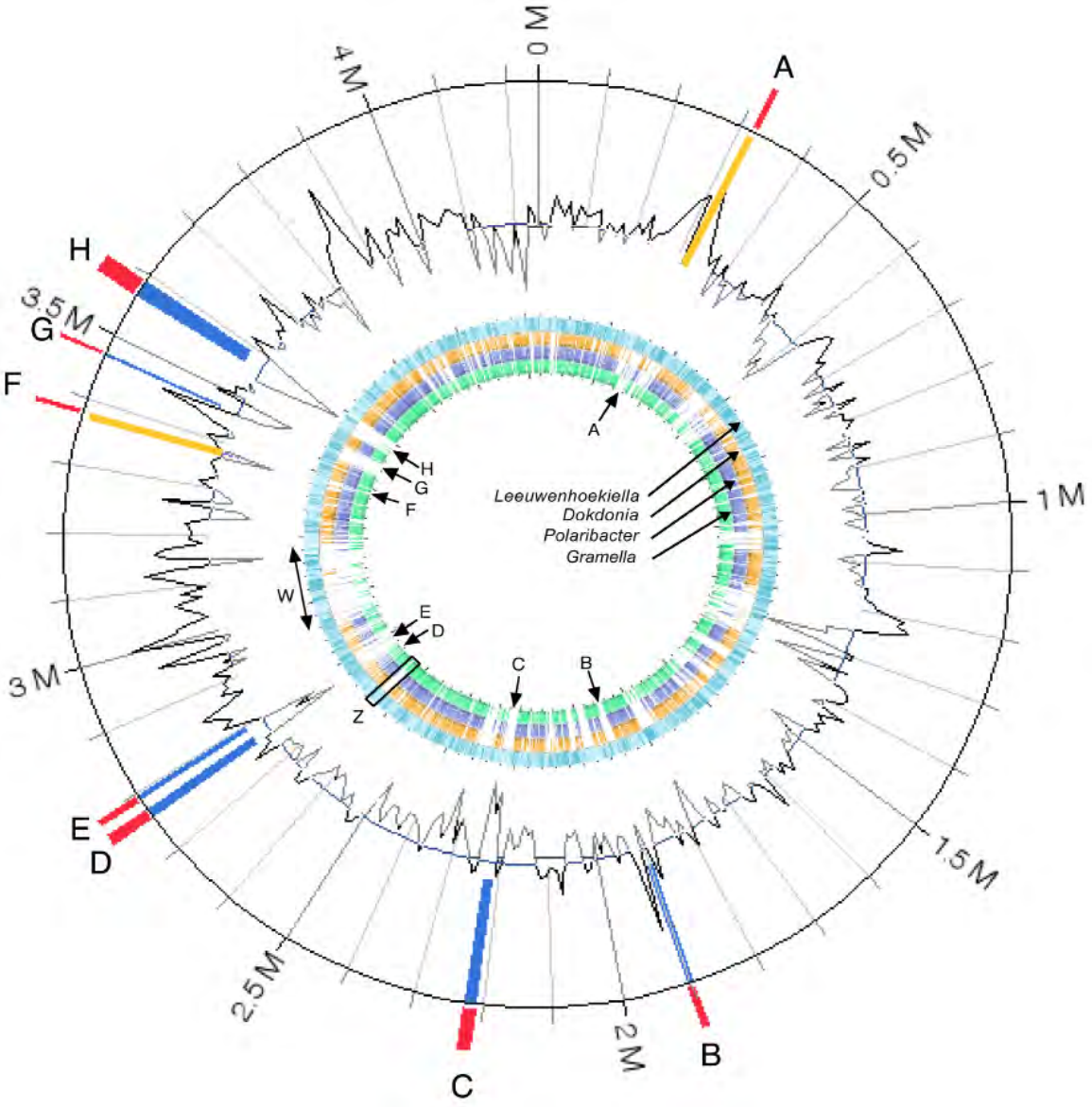


Figure 2. The blue ring shows the genome of *Leeuwenhoekiella blandensis* MED217 ordered from 0 to 4.24 Mb. The next outer ring shows the G+C content along the genome. The outermost ring shows the areas identified as genomic islands with the IslandViewer software. Red, blue, and yellow indicate the three different tools used for island prediction in this package. The inner colored rings are the genomes of *Dokdonia* sp. MED134, *Polaribacter* sp. MED152, and *Gramella forsetii* compared to that of MED217. Each one of these genomes has been compared to that of MED217 by reciprocal best matches and, whenever a match was found, the gene of these genomes was placed next to the position of MED217 best match. That is, these three genomes do not keep their original topology so that genomes can be compared gene to gene. The letters A to H indicate genes present in MED217 and absent from some or all the other genomes, that were identified as genomic islands. Rectangle labeled Z and double arrow labeled W show areas in MED217 that were absent from some of the other genomes, but were not genomic islands.

three MED strains, MED217 could use arabinose while the two other bacteria could not (O. I. Nedashkovskaya, personal communication).

The region labeled W is quite extensive and most of it is missing in the other three genomes. This region contained approximately 218 ORFs, about half of these were hypothetical proteins and 25 more were only identified putatively. Among the remaining ORFs the number of genes involved in sugar metabolism was remarkable. For example, four copies of beta-galactosidase, one arabinosidase and several regulatory proteins including two of the arabinose operon were found. This suggests that MED217 is capable of utilizing a large number of sugars that the other three are not.

The lack of genomic islands, the absence of certain metabolic pathways, and the reduced number of paralogous families and genes within such families, are three important mechanisms accounting for the reduced size of the genomes of the PR+ Bacteroidetes.

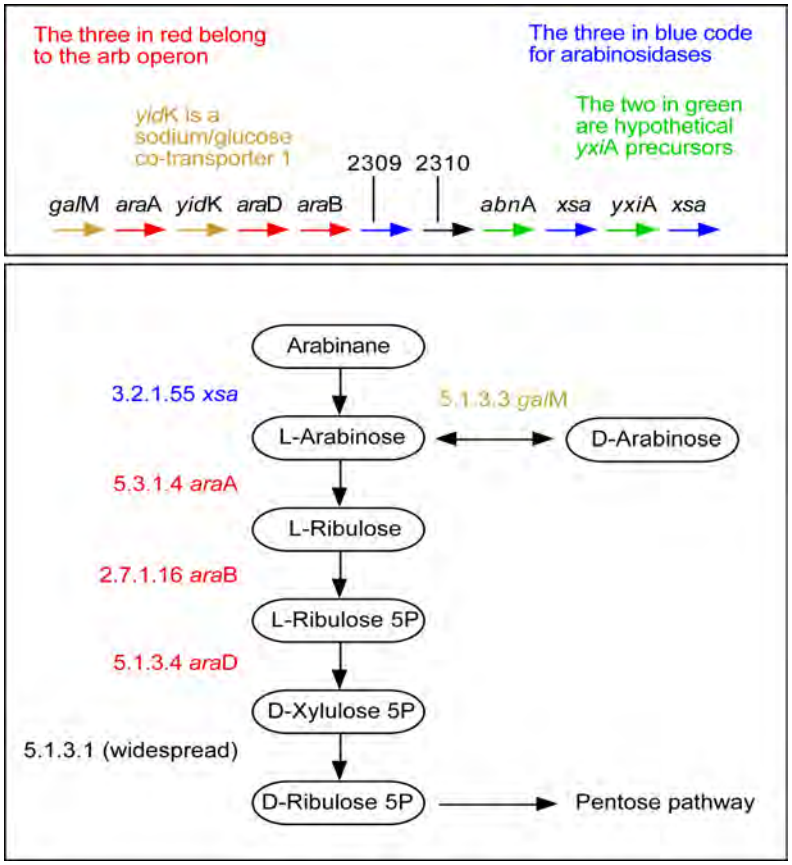


Figure 3. Genes identified in section Z of the *Leeuwenhoekiella blandensis* MED217 genome. They include the structural genes of the arabinose operon (red), plus other genes involved in sugar metabolism (blue and green) and transport (*yid*K).

Whether the PR+ genomes have experienced streamlining or whether, on the contrary, the PR- genomes have increased in size cannot be determined. However, it is remarkable that all the Bacteroidetes PR+ genomes are small. The percentage of genes in paralogous families followed the well-known (Pushker *et al.,* 2004; Woyke *et al.,* 2009) linear relationship with genome size for a large selection of bacteria (small dots corresponding to "other" bacteria in Fig. 1). The seven PR+ Bacteroidetes (yellow triangles) had genomes smaller than the around 50 PR- Bacteroidetes (pink triangles). Among the eight PR+ Proteobacteria (pink circles), only the *Pelagibacter*-like genomes were smaller than those of the PR+ Bacteroidetes, while other PR+ Proteobacteria (e.g Gammaproteobacteria) had larger genomes. The number of known PR+ marine bacteria is still very low and this pattern needs to be confirmed with more genomes. However, the environmental Bacteroidetes genomes MS024-2A and MS024-3C also followed this trend (see the two smallest genomes labeled as PR+ Bacteroidetes in Fig. 1): they had the PR gene and their genomes were estimated to be smaller than that of *Polaribacter* sp. MED152, adding strength to the argument of PR+ Bacteroidetes having small genomes.

It is tempting to conclude that possession of PR allows bacteria to reduce their genomes. Perhaps the extra mechanism for energy conservation allows the cells to be less versatile in their carbon source preferences. In other words, they may not need to carry the genes for many different carbon utilization pathways because light energy makes them more independent from organic compounds as energy sources. The small number of transporters is consistent with this idea.

Unlike the PR+ Bacteroidetes, the PR+ Gammaproteobacteria varied more widely in their genome sizes (Fig. 1). It has been suggested that PR plays several roles in different marine bacteria (Fuhrman *et al.,* 2008; DeLong and Béjà 2010). So far, higher cell yields in the light than in the dark have been demonstrated only in *Dokdonia* sp. MED134 (Gómez-Consarnau *et al.,* 2007; Kimura *et al.,* 2011) and in *Polaribacter* sp. MED152 (Férnandez-Gómez *et al.,* in preparation). Enhanced growth in the light has not been seen in *Pelagibacter* (Giovannoni *et al.,* 2005) or in Gammaproteobacteria (Stingl *et al.,* 2007). On the other hand, the usefulness of PR under starvation conditions has been demonstrated in one Gammaproteobacteria (Gómez-Consarnau *et al.,* 2010) and in *Pelagibacter* (Steindler *et al.,* 2011). So far, the two Bacteroidetes mentioned are the only bacteria that have been found to show higher growth yields more in the light than in the dark under low carbon concentration conditions.

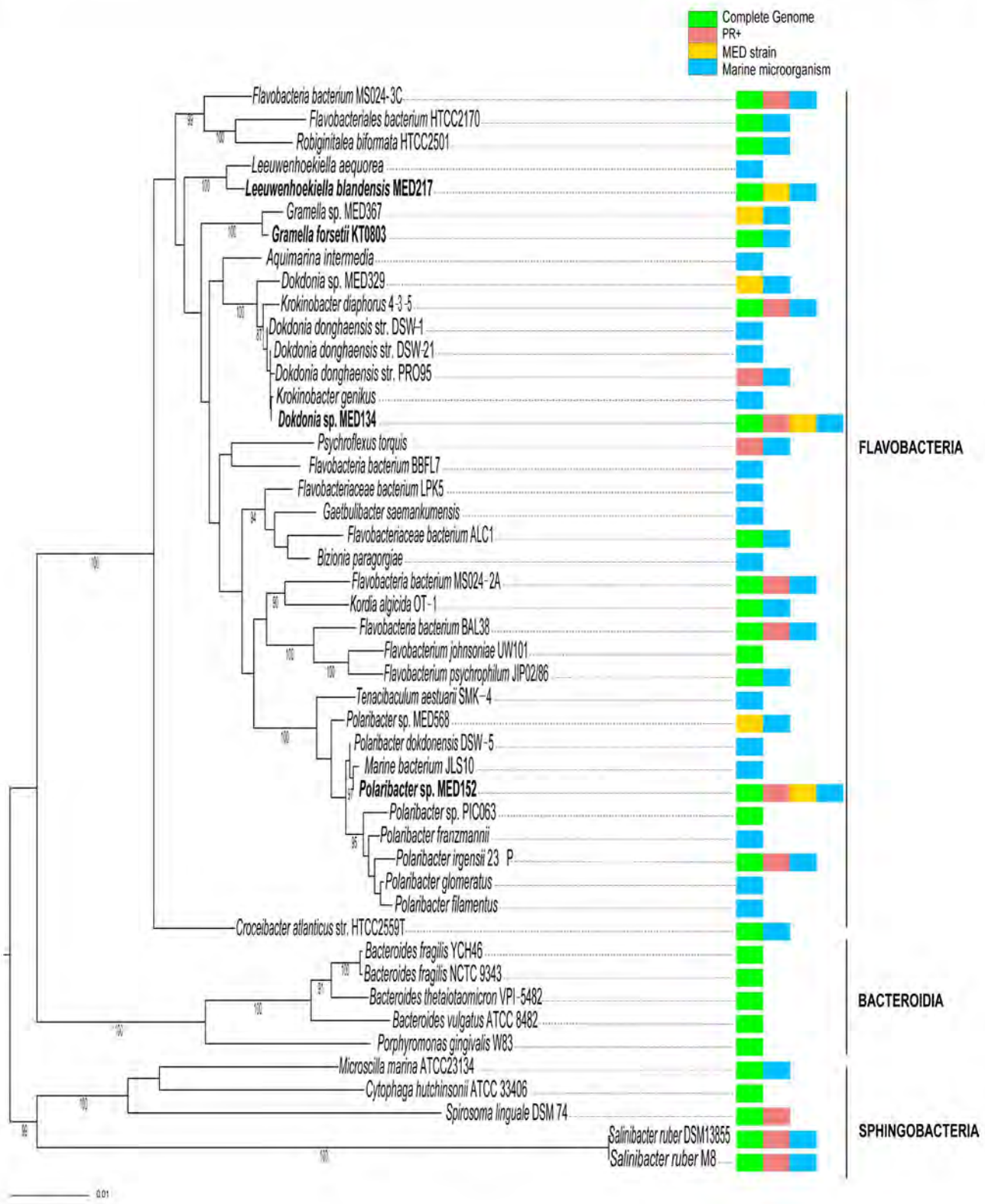


Figure 4. Phylogenetic tree of the Bacteroidetes emphasizing the marine bacteria (blue), those with the whole genome sequenced (green), with PR (red) and those belonging to the culture collection from the Blanes Bay Microbial Observatory (yellow).

2. PR as an adaptation to life in the oligotrophic surface ocean

In *Polaribacter* sp. MED152, it was proposed that PR, in combination with a large suite of light sensing genes and a remarkable number of proteins involved in anaplerotic carbon fixation, represented adaptations to survive in the surface ocean while traveling between particles (González *et al.,* 2008). This bacterium has a relatively small genome and it was postulated that it was only able to use a small number of monomeric carbon compounds, a fact consistent with the very low number of transporters compared to other bacteria and with a lower number of sugars used compared to MED217 (O. Nedashkovskaya pers. comm.).

The presence of PR in the Bacteroidetes with the smaller genomes and its absence in those with larger genomes suggests that this strategy may be common to the other PR+ Bacteroidetes. However, the possession of PR is a characteristic widely distributed throughout the different clusters of Flavobacteria (Fig. 4). Moreover, both PR+ and PR- bacteria are found within the same clusters, such as the flavobacterium BAL38 (PR+) and *Flavobacterium psychrophilum* JIP02/86 (PR-) for example. In this section we examine which of the remaining features of *Polaribacter* sp. MED152 are consistent with the presence or absence of PR in other Flavobacteria.

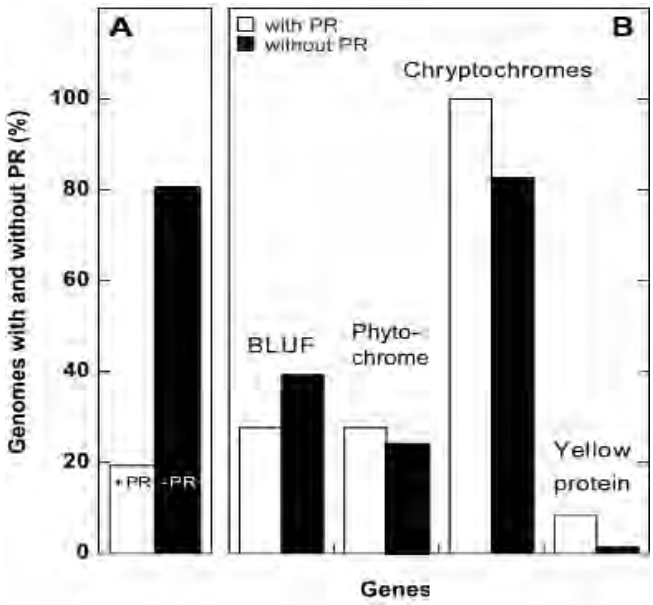


Figure 5. A. Percent of genomes from a collection of 185 marine bacteria with proteorhodopsin (PR+) and without it (PR-). **B.** Percent of PR+ or PR- genomes with different light sensing domains.

First, lets consider the distribution of light sensing domains in bacteria in general. Figure 5 shows a comparison between bacteria with and without PR. All marine bacterial genomes sequenced are included (185 genomes), providing a good representation of the main marine groups of bacteria available in pure culture. Close to 20% of this collection had the PR gene (Fig. 5A). The other light related domains were present in different numbers of genomes. Thus, cryptochromes were widespread both in PR+ and PR- bacteria, while the photoactive yellow protein (PYP) was present in very few genomes (Fig. 5B). In the case of the PYP a higher percent of PR+ bacteria possessed the domain, but the absolute number of genomes involved was so low that the difference cannot be considered significant. In the case of cryptochromes all the PR+ bacteria had them and 83% the PR- cells. Again, this difference is not very large. In conclusion, no significant difference in the number of light sensing domains seems to exist between bacteria with and without PR.

Table 3. Bacteria in which a PR coding gene has been identified and distribution of paticular light-sensing domains (a black square means presence, a minu sign means absence).

Organism	BLUF	Pchr	AnCchr	DASH	Photoly
Bacteroidetes					
Polaribacter sp. MED152	■	■	■	■	■
Polaribacter irgensii 23-P	-	-	■■	■	■
Dokdonia sp. MED134	■	■	■	■	■
F. bacterium BAL38	-	-	■	■	■
Alphaproteobacteria					
P. ubique HTCC1002	-	-	-	-	■
P. ubique HTCC1062	-	-	-	-	■
Octadecabacter antarcticus 307	-	-	■	-	■
Betaproteobacteria					
Methylophilales HTCC2181	-	-	-	■	■
Gammaproteobacteria					
Vibrio harveyi ATCC BAA-1116	-	-	■	■	■
Vibrio angustum S14	-	-	■	■	■
HTCC2143	-	-	■	■	■
HTCC2207	-	-	-	■	■
Photobacterium sp. SKA34	-	-	■	■	■
Marinobacter sp. ELB17	-	-	■■	■	■

BLUF: Blue-light sensing using flavin; **Pchr:** Phytochrome; **AnCchr:** Animal cryptochrome and (6-4) photolyase family; **DASH:** Cryptochrome DASH family; **Photoly:** Photolyase class I.

When looking at the distribution of these light sensing domains within PR containing bacteria, however, different patterns emerge (Table 3). All of them possess the photo-lyase class I gene that codes for a widely distributed DNA-repair protein. Two other types of this family of proteins are present in most Gammaproteobacteria and in all the Bacteroidetes. However, they are missing from the two known *Pelagibacter* genomes. The DASH family is common to most Bacteroidetes and is, therefore, not specific of the PR+. It is also found in Cyanobacteria, but not in Proteobacteria. However, the animal cryptochrome-like family is only found in the PR+ Bacteroidetes and in some Gammaproteobacteria. Finally, two of the four Bacteroidetes have proteins with BLUF and phytochrome domains. This comparison

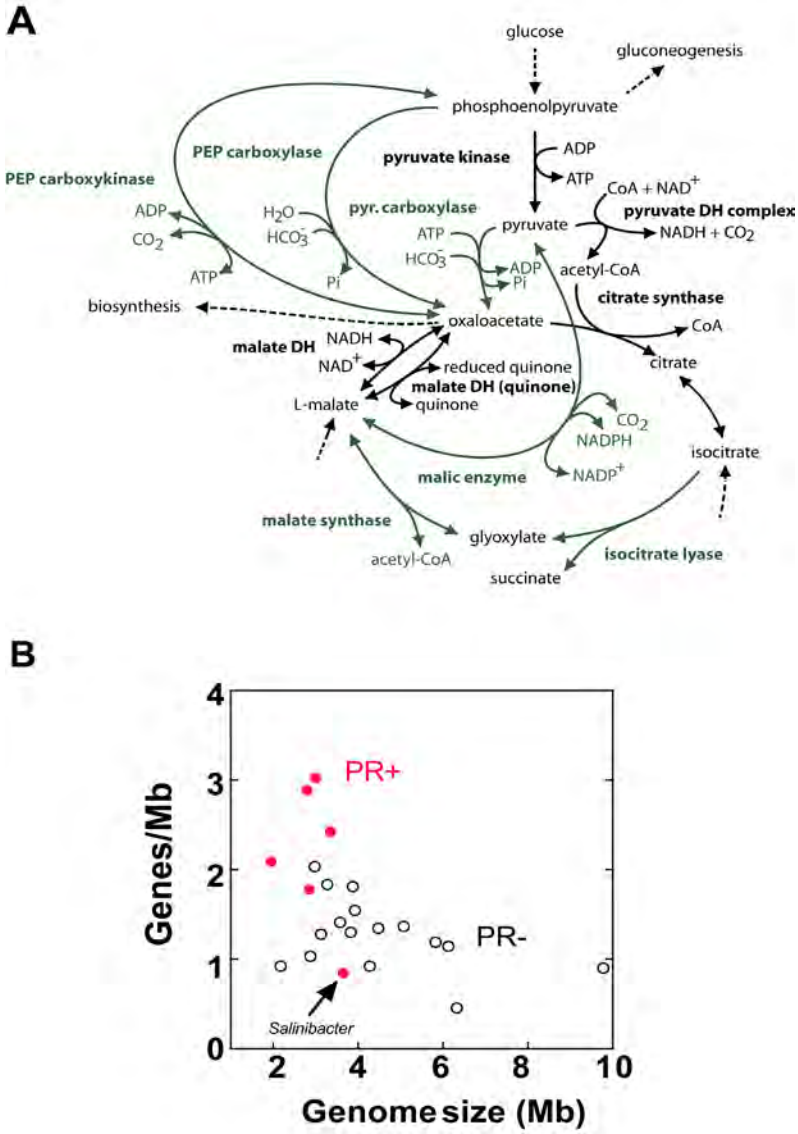


Figure 6. A. Anaplerotic metabolism indicating the enzymes involved in CO_2 fixation. **B.** Genes involved in anaplerotic fixation and transport of CO_2 in different genomes versus genome size. PR+ genomes are in pink.

indicates that different PR+ bacteria must use light with different strategies. It is intriguing that even two very closely related bacteria such as the two *Polaribacter* strains have different sensors, while bacteria belonging to two different genera, like *Polaribacter* and *Dokdonia*, do share the same light sensing domains. This fact stresses that care must be taken when extrapolating from studies of one bacterium to others, especially closely related ones. At any rate, *Polaribacter* sp. MED152 and *Dokdonia* sp. MED134 share the same large number of light sensing domains, suggesting that they may use similar strategies for light utilization.

The second feature hypothesized to enhance survival under oligotrophic conditions was the large number of proteins involved in anaplerotic carbon fixation found in MED152 (González *et al.,* 2008), which might be a characteristic feature in all PR+ Bacteroidetes.
We consider several proteins in this group (Fig. 6A): a) two enzymes involved in the glyoxylate shunt (isocitrate lyase and malate synthase), b) two enzymes involved in CO_2 fixation with phosphoenolpyruvate (PEP) (PEP carboxylase, PEP carboxykinase), c) two enzymes that fix CO_2 with pyruvate (pyruvate carboxylase and malic enzyme), and d) three proteins involved in bicarbonate acquisition and interconversion (carbonic anhydrase and the two bicarbonate transporters *bic*A and *sbt*A).

PR+ Bacteroidetes had small genomes with a large number of genes involved in these three processes (Fig. 6B). The other Bacteroidetes had variable and, in most cases, lower numbers. All the bacteria having more than two such genes per Mb, also possessed PR. These included *Pelagibacter ubique* and the Gammaproteobacteria HTCC2207 (not shown in the figure). Most other bacteria had a value lower than 1.7. The only PR+ bacterium with values significantly lower than 2 was *Salinibacter ruber*. The latter, however, is a specialist of hypersaline environments and its life strategy is likey very different from those of marine bacteria. Thus, the possesion of a large suite of genes involved in anaplerotic carbon fixation seems to be linked with the possession of PR in Bacteroidetes (present in five out of six genomes). This, however, does not mean that other bacteria do not have anaplerotic metabolism, only that they do not have as many enzymes capable of fixing CO_2.

The third point of the model was that *Polaribacter* sp. MED152 was limited in the number of soluble compounds it could use for growth. This feature was reflected in a low number of transporters relative to other bacteria. The number of transporters identified in different bacteria is shown against genome size in Fig. 7A. It is striking that the Bacteroidetes clearly follow a pattern different from that of the other bacteria. Significant regressions can be calculated separately for Bacteroidetes and the other bacteria with slopes of 34.6 and 56.5 transporters per Mb respectively. Since there are marine and non marine bacteria in

both groups, this feature is clearly a characteristic of Bacteroidetes in general and does not represent an adaptation to marine life (Fig. 7B). Rather, marine Bacteroidetes must cope with this relatively low number of transporters when living in the ocean.

In summary, there seems to be a co-occurrence between the presence of the PR gene and a large number of anaplerotic CO_2 fixation genes in the genomes of Bacteroidetes and Proteobacteria. In addition, *Polaribacter* sp. MED152 and *Dokdonia* sp. MED134 share the

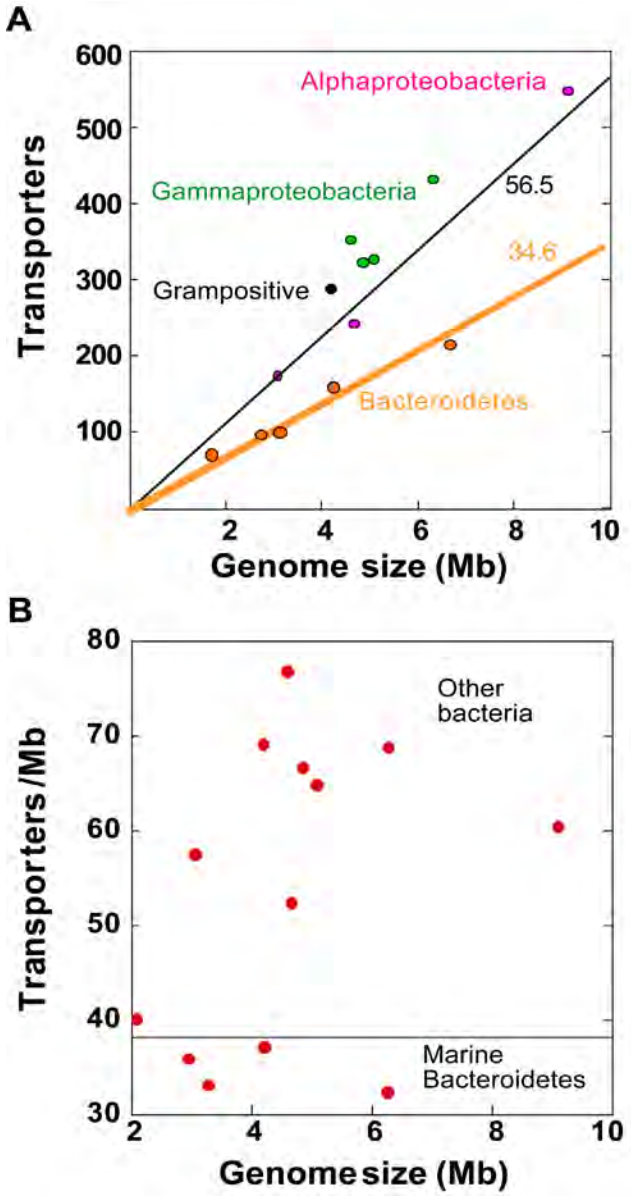


Figure 7. A. Number of transporters versus genome size for Bacteroidetes (orange symbols) and other bacteria. Numbers are the slopes of linear regressions fitted to the data. **B.** Number of transporters per megabase versus genome size. All marine Flavobacteria fall below the line of 38 transporters per Mb.

traits of having small genomes, and possessing a large number of light sensing/utilization genes, including three photolyase related proteins. *Pelagibacter* strains share with the Bacteroidetes the small genome and the number of anaplerotic CO_2 fixation genes, but not the light sensing and utilizing genes. Therefore, there must be at least three different strategies to use PR in marine bacteria (employed by Bacteroidetes, *Pelagibacter*-like and Gammaproteobacteria respectively).

3. Adaptation to polymer degradation on particles

The main lifestyle of marine Bacteroidetes is assumed to be attachment to particles and degradation of polymers. This assumption is based on a few analyses of free-living and attached bacterial diversity (DeLong *et al.,* 1993), some microautoradiography experiments (Cottrell and Kirchman 2000), a correlation of Bacteroidetes abundance with phytoplankton blooms (Pinhassi *et al.,* 2004), microcosm experiments in which Bacteroidetes were enriched on organic matter particles (Pedrotti *et al*., 2009), and microscopic observations of CARD-FISH stained Bacteroidetes associated to the phycosphere of nanoplankton cells in the Atlantic Ocean (Gómez-Pereira *et al*., 2010). The genomes of both *G. forsetii* and *Polaribacter* MED152 were in accordance with this notion. *Polaribacter*, for example, has a large number of genes enconding for peptidases and glycoside hydrolases, necessary to degrade polymers. It also had the genes for gliding motility, a motility mechanism that can only be used on a surface, and it also had many proteins involved in adhesion (González *et al.,* 2008). In this section we explore whether this characteristics can be extended to the marine Bacteroidetes in general and whether differences can be found between PR+ and PR- bacteria.

Adhesion and Gliding

The four marine Flavobacteria examined in more detail had between 21 and 27 domains involved in surface adhesion (Supplementary Table 7). When normalized for genome size, this represented at least 6 adhesion related genes per Mb (Fig. 8A). In contrast, other pelagic bacteria such as the Alphaproteobacteria *Pelagibacter ubique* and *Ruegeria pomeroyi,* and the Gammaproteobacteria *Idiomarina loihiensis* and *Vibrio parahaemolyticus* only had one or two genes per Mb. As could be expected, *Bacteroides thetaiotaomicron* had the most adhesion genes per Mb of all (nine), since it lives attached to the intestinal epithelium of vertebrates Fig. 8).

The marine Bacteroidetes also had two to three times more glycosyl transferases per Mb than the Proteobacteria (with the exception of *P. ubique*, Fig. 8B). These proteins are usually positioned in the outer membrane and generate polysaccharides for e.g., attachment. Altogether, this clearly indicates that attachment to particles is an important feature of marine Flavobacteria in general (independently of possession or absence of PR), while this is not the case for common planktonic Proteobacteria such as *Pelagibacter*, *Ruegeria,* or *Vibrio*.

Flagella are not commonly used to move on surfaces (Jarrell and McBride 2008). Most motile bacteria living on surfaces use gliding motility, a name that actually includes several different molecular mechanisms of movement on surfaces (Jarrell and McBride 2008). The four marine Flavobacteria had the full complement of 15 genes for gliding motility. Moreover, the three MED strains were shown to have gliding motility under the microscope (O. I. Nedashkovskaya, personal communication). Again this is a characteristic common to many marine Flavobacteria and it is not specific of PR+ species.

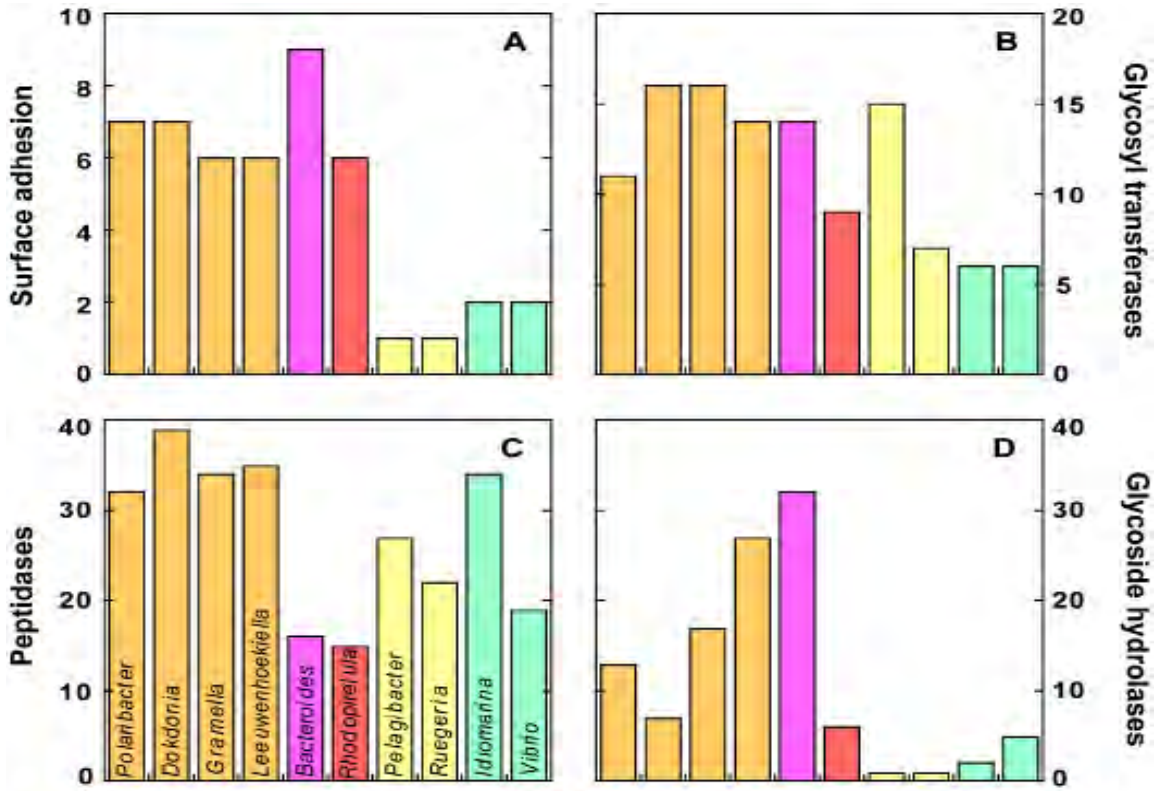


Figure 8. Numbers of different enzymes per megabase of genome for a selection of bacteria: Marine Bacteroidetes (orange) *Polaribacter* sp. MED152, *Dokdonia* sp. MED134, *Gramella forsetii*, and *Leeuwenhoekiella blandensis* MED217; *Bacteroides thetaiotaomicron* (pink); *Rhodopirellula baltica* (maroon); the Alphaproteobacteria (yellow) *Pelagibacter ubique* and *Ruegeria pomeroyi*; and the Gammaproteobacteria (green) *Idiomarina loihiensis* and *Vibrio parahaemolyticus*.

Polymer degrading enzymes: Peptidases and Glycoside hydrolases

Our analysis showed that, in general, the number of peptidases and glycoside hydrolases (GH) increased with the size of the genome in all bacteria (Fig. 9A and C). Most Bacteroidetes, however, tended to have more of these enzymes than the average bacterium, irrespectively of the genome size (Fig. 9). This is one of the major observations that indicates the dedicated role of marine Bacteroidetes as polymer degraders. Marine Flavobacteria had more GHs per Mb than other planktonic bacteria (Fig. 9D). The difference was statistically significant. *Dokdonia* sp. MED134 had a lower number of GHs, similar to those of *V. parahaemolyticus* and *R. baltica*, but still higher than those of other planktonic bacteria such as *Pelagibacter*, *Ruegeria* or *Idiomarina* (Fig. 8D and 9D). *B. thetaiotaomicron* had the largest number of GHs per Mb in accordance with its specialization on polysaccharide degradation in the

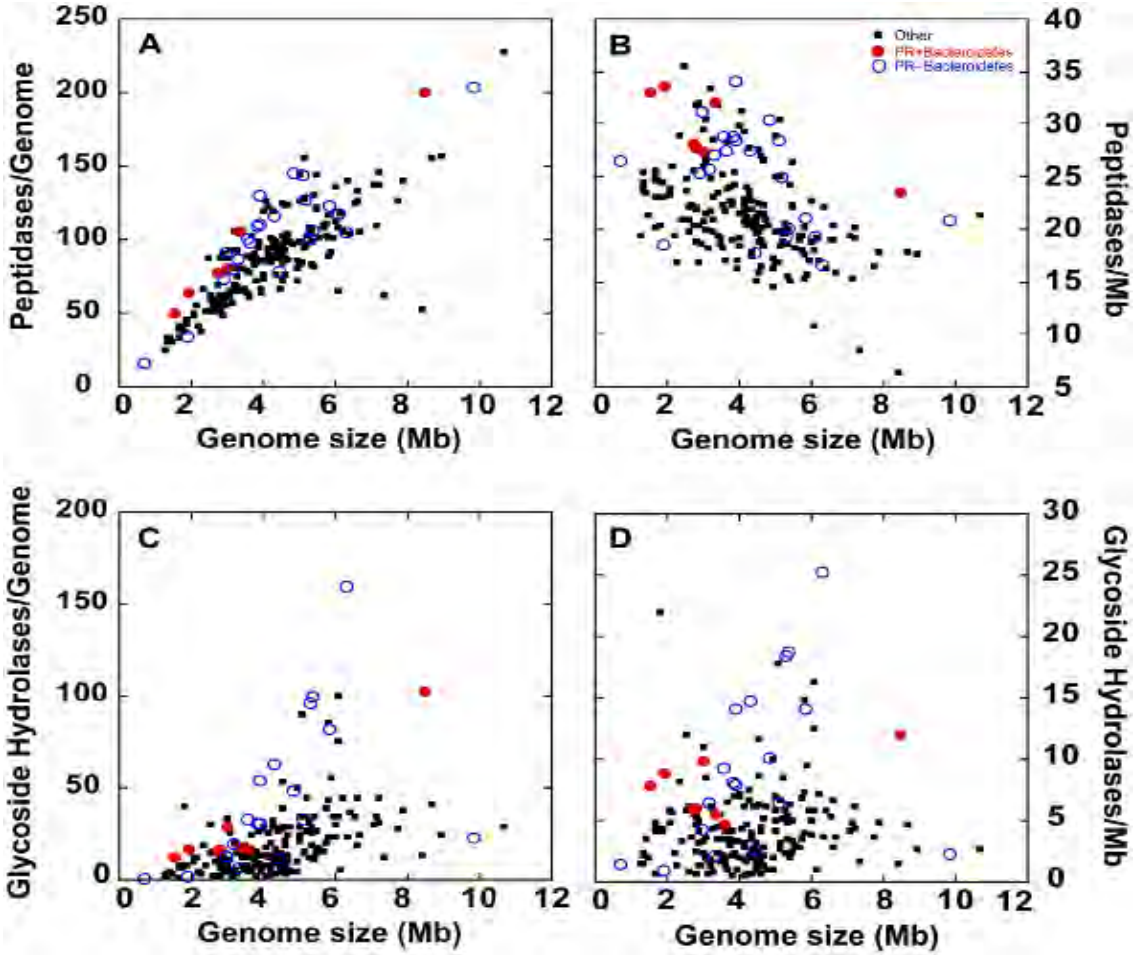


Figure 9. A. Number of peptidases versus genome size. **B**. Peptidases per Mb of genome. **C.** Glycoside hydrolases versus genome size. **D**. Glycoside hydrolases per Mb. PR+ Bacteroidetes are in purple, PR- Bacteroidetes in blue.

vertebrate digestive tract. In the case of peptidases, the marine Flavobacteria had the largest numbers of genes per Mb (Fig. 8C and 9B). Again, the difference between Bacteroidetes and other marine bacteria was also significant. Moreover, the PR+ Bacteroidetes had significantly larger numbers of peptidases per Mb than PR- Bacteroidetes. As expected, *B. thetaiotaomicron* had a lower number of peptidases, since it is a polysaccaride specialist (Martens *et al.,* 2009). The relatively high number of peptidases per Mb in *P. ubique* is remarkable, especially combined with its low number of GHs, suggesting a preference for proteins over polysaccharides. Finally, the gammaproteobacterium *Idiomarina loihiensis* also had a large number of peptidases per Mb, in accordance with its putative specialization in the fermentation of aminoacids and proteins (Hou *et al.,* 2004).

In summary, marine Flavobacteria had a relatively large complement of both GHs and peptidases compared to other marine bacteria and, moreover, they had as many peptidases as other protein specialists.

Another striking observation was that marine Bacteroidetes had many more peptidases than GHs (Fig. 10). This was not the case for the non-marine Bacteroidetes examined. This strongly suggests a specialization of marine bacteroidetes on the degradation of proteins. This is consistent with the experimental studies using microautoradiography, in which a preference for proteins has been demonstrated (Cottrell and Kirchman 2000). Gómez-Pereira *et al*. (2011) analyzed Bacteroidetes fosmids from two regions in the North Atlantic Ocean. They found many polysaccharide degrading enzymes in the phytoplankton rich polar waters and an abundance of protein degrading enzymes in the subtropical North Atlantic. More isolates from polar waters will have to be examined to see if the relationship in Figure 6 holds for all marine Bacteroidetes or only for those in temperate regions.

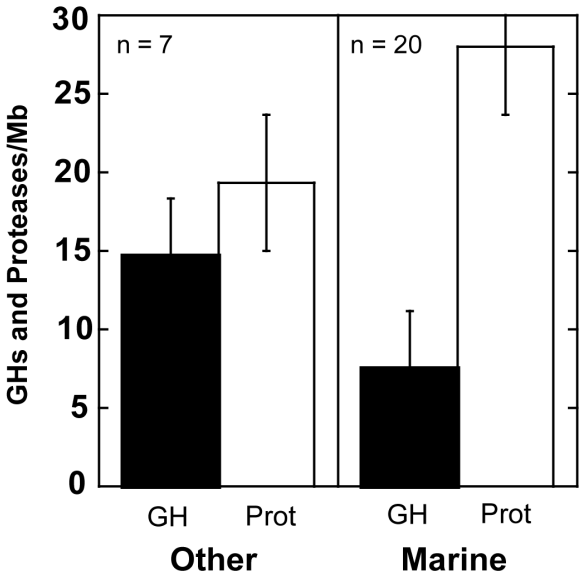


Figure 10. Peptidases or glycoside hydrolases per Mb for marine and non-marine Baceroidetes.

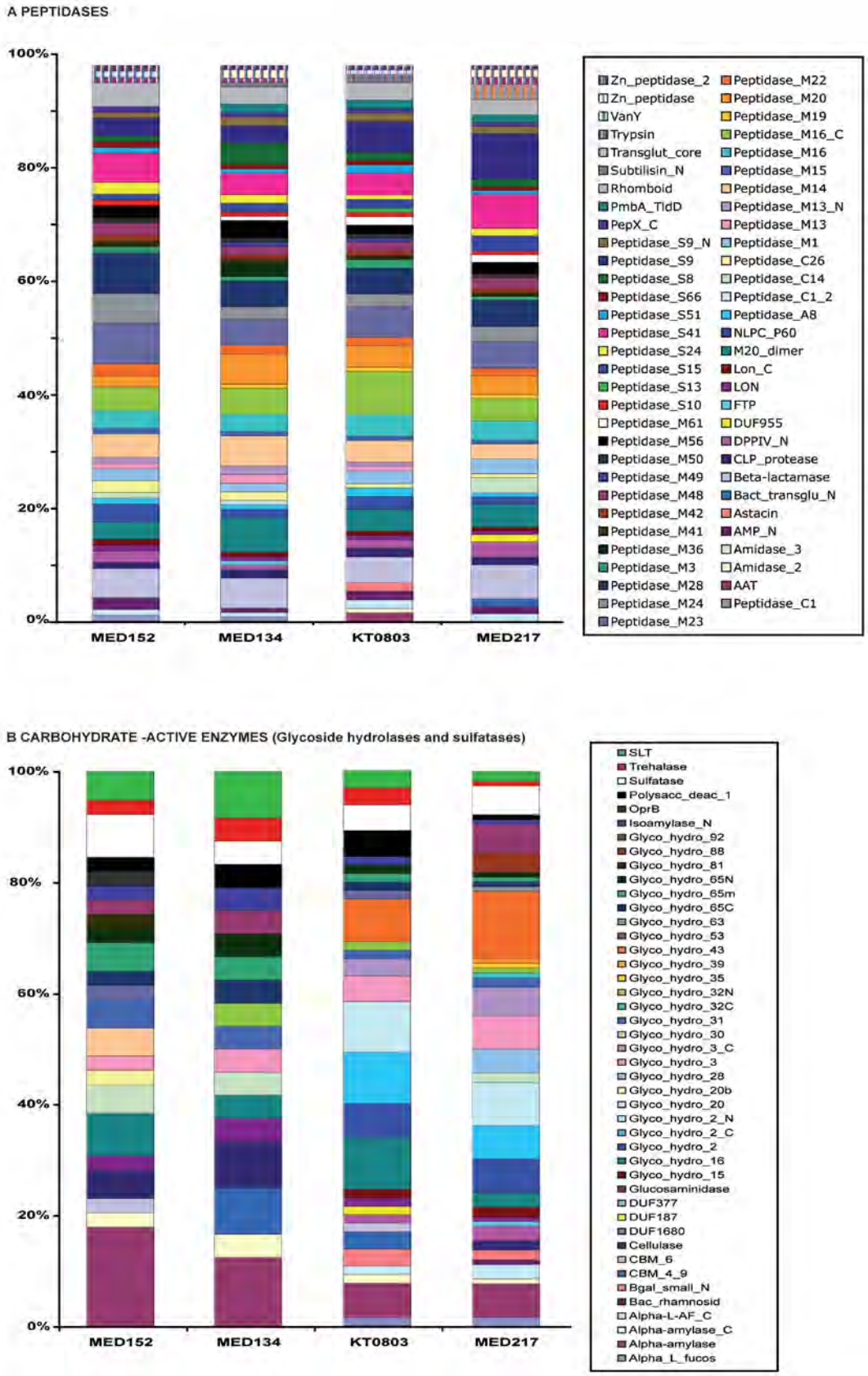


Figure 11. Diversity of peptidases **(A)** and glycoside hydrolases and sulfatases **(B)** in the four marine Bacteroidetes analyzed.

Despite these general characteristics, each one of the four bacteria showed a different suite of GHs and peptidases (Fig. 11). This indicates variations on a shared theme. Probably, in combination with other genes, these differences allow the various species to occupy slightly different niches. We calculated diversity indices for these enzymes and found that the four bacteria had very similar values. In the case of peptidases the indices varied between 3.57 and 3.68 and in the case of GHs between 2.82 and 3.20. These indices show that not only do these bacteria have more peptidases than GHs, but that there is a larger diversity of the former. Thus, the conclusion that protein degradation is the main specialty of marine Bacteroidetes is robust.

4. Families of paralogous genes

A fruitful approach to ascertain the principal way of life of bacteria is to look at the main families of paralogous genes (Suppl. Table 8). The three largest families were the same for the four Flavobacteria: Two-component systems (between 30 and 54 genes), TonB-dependent receptors (18-50), and ABC transporters (22-30). These numbers are in striking contrast to those of two common Proteobacteria such as *E. coli* or *Ruegeria pomeroyi*. The latter bacteria had ABC transporters as the largest family, followed by LysR one-component systems, and two-component systems in the third place. Moreover, the Bacteroidetes had twice the number of TonB dependent receptors per Mb than the Proteobacteria. These large differences are surely related to the different life styles of these bacteria.

The Bacteroidetes had between one half and one third of the ABC transporters per Mb than Proteobacteria. This lower number of ABC transporters is in line with the generally low number of transporters for low molecular weight compouds that were found in the genomes of both *Polaribacter* and *Dokdonia*. This feature is now seen to be general to other marine Bacteroidetes.

Finally, the abundance of TonB dependent receptors also suggests a specialization in degradation of polymers. In particular, the four bacteria have between 6 and 17 *susC* plus *susD*-like pairs of genes. In *Bacteroides thetaiotaomicron*, SusC is a member of the TonB receptor family specialized in the transport of oligosaccharides from the outer membrane into the periplasmic space. SusD, in turn, is necessary to bind polysaccharides to the outer cell membrane. These two genes alone are enough to account for 60% of the polysaccharide degrading ability of *B. thetaiotaomicron* (Martens *et al.,* 2009). This is likely the role of the many *susC* plus *susD* pairs in the four Bacteroidetes. Moreover, these pairs are always next to genes encoding polymer degrading enzymes: sulfatases, amylases, glycoside hydrolases,

peptidases, and alkaline phosphatases among other enzymes (Fig. 12). In *B. thetaiotaomicron* there are 101 individual pairs of "*susC*-like" and "*susD*-like" genes, with the former always positioned immediately upstream of the latter. This bacterium is a polysaccharide specialist of the distal digestive tract of mammals and has a large genome of 6.26 Mb (resulting in 16 *sus*C-*sus*D pairs per Mb). In *Flavobacterium johnsoniae*, a soil organism specialized in degradation of polysaccharides, there are 50 TonB dependent receptors and 10 of these also have the *sus*C plus *sus*D combination (McBride *et al.,* 2009). This represents only 1.7 pairs per Mb. In the four marine Bacteroidetes, with much smaller genomes, we found 5, 4, 17, and 21 such pairs in *Polaribacter*, *Dokdonia*, *Leeuwenhoekiella*, and *Gramella* respectively (1.7, 1.2, 4.5 and 5 per Mb respectively) (Suppl. Table 9). This is in line with the hypothesis of an adaptation to the use of polymers in the marine Bacteroidetes.

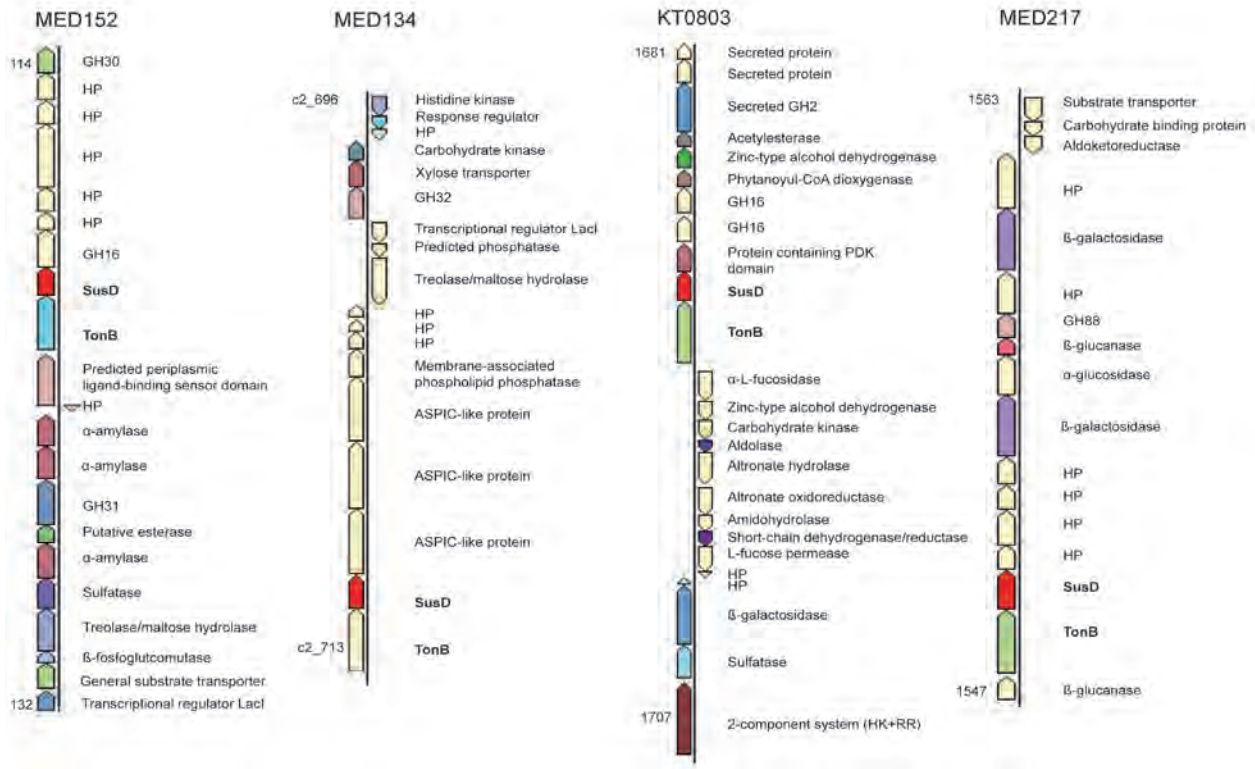


Figure 12. Clusters of genes putatively involved in the attachment and degradation of polymeric compounds and containing TonB dependent/ligand-gated channel genes and SusD. One example from each of the four bacteria is shown. The complete list can be found in Suppl. Table 9. **GH:** Glycoside hydrolase; **HP:** Hypothetical protein; **HK:** Histidine kinase; **RR:** Response regulator.

CONCLUSIONS

Analysis of the *Polaribacter* sp. MED152 genome suggested a dual life style (González *et al.,* 2008): a) growth attached to particles degrading poymers and b) dispersal through the water column obtaining energy from light through proteorhodopsin. Each one of these two states was supported by a series of functions. The attached phase would require genes involved in adhesion, glycosyl transferases, gliding motility, a large number of peptidases and glycoside hydrolases for degradation of polymers, and SusC-SusD pairs for attachment and degradation of polymers. The free-living phase, in turn, would rely on proteorhodopsin, a large number of light sensing proteins, and an enhanced anaplerotic fixation of CO_2.

Our comparative study has shown that this whole complement of genes is shared with *Dokdonia* sp. MED134 but, interestingly, not by the close relative *Polaribacter irgensii* 23-P. The two characteristics shared by all known PR+ Bacteroidetes are small genome size and a higher number of genes involved in CO_2 fixation per Mb than the PR- Bacteroidetes. Within the PR+ Proteobacteria, analysis of their genomes identified at least two different strategies corresponding to *Pelagibacter*-like bacteria on the one hand, and Gammaproteobacteria on the other.

Independently of the presence or absence of PR, the marine Bacteroidetes share the capacity for adhesion and gliding motility, the presence of abundant glycosyl transferases, a large number of polymer degrading enzymes including glycoside hydrolases and, especially, peptidases, and a large relatively number of SusC-SusD pairs associated to many different degrading enzymes. This confirms the role of this abundant group of marine bacteria as degraders of particulate matter, especially of proteins.

ACKNOWLEDGEMENTS

B. F.-G. was a recipient of a I3P grant from CSIC. The genomes of the three MED strains were sequenced by the JCVI through the Marine Microbial Initiative of the Gordon and Betty Moore Foundation. We thank Thierry Lombardot for bioinformatic analysis. This work was supported by Grant “GEMMA” CTM2007-63753-C02-01/MAR from the Spanish Ministerio de Ciencia e Innovación.

REFERENCES

Alonso C, Warnecke F, Amann R, Pernthaler J. (2007). High local and global diversity of Flavobacteria in marine plankton. *Environ Microbiol* **9:** 1253-1266.

Alonso-Sáez L, Gasol J. (2007). Seasonal variations in the contributions of different bacterial groups to the uptake of low-molecular-weight compounds in Northwestern Mediterranean Coastal waters. *Appl Environ Microbiol* **73:** 3528-3535.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* **215:** 403-410.

Amaral-Zettler L, Artigas LF, Baross J, P.A. LB, Boetius A, Chandramohan D *et al.,* (2010). A Global Census of Marine Life. In: Alasdair D. McIntyre, Ed. Life in the World's Oceans. Blackwell Publishing Ltd., pp 223-245.

Andersson DI, Hughes D. (2009). Gene amplification and adaptative evolution in bacteria. Annu Rev Genet **43**:167-195.

Bauer M, Kube M, Teeling H *et al.,* (2006). Whole genome analysis of the marine Bacteroidetes *Gramella forsetii* reveals adaptations to degradation of polymeric organic matter. *Environ Microbiol* **8:** 2201-2213.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340: 783-795.

Castresana J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17:** 540-552.

Cottrell MT, Kirchman DL. (2000). Natural assemblages of marine proteobacteria and members of the Cytophaga-Flavobacteria Cluster consuming Low- and High-Molecular-Weight dissolved organic matter. *Appl Environ Microbiol* **66:** 1692-1697.

DeLong EF, Franks DG, Alldredge AL. (1993). Phylogenetic diversity of aggregate-attached vs. free-fiving marine macterial assemblages. *Limnol Oceanogr* **38:** 924-934.

DeLong EF, Béjà O. (2010). The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times. *PLoS Biol* **8:** e1000359.

Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz H-R *et al.,* (2008). The Pfam protein families database. *Nucleic Acids Res* **36:** D281-D288.

Fuhrman JA, Schwalbach MS, Stingl (2008). Proteorhodopsins: an array of physiological roles? *Nat Rev Micro* **6:** 488-494.

Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D *et al.,* (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309:** 1242-1245.

Glöckner FO, Fuchs BM, Amann R. (1999). Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence in situ hybridization. *Appl Environ Microbiol* **65:** 3721-3726.

Goldberg SMD, Johnson J, Busam D, Feldblyum T, Ferriera S, Friedman R *et al.,* (2006). A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. *Proc Natl Acad Sci USA* **103**:11240-11245.

Gómez-Consarnau L, González JM, Coll-Lladó M, Gourdon P, Pascher T, Neutze R *et al.,* (2007). Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. *Nature* **445:** 210-213.

Gómez-Consarnau L, Akram N, Lindell K, Pedersen A, Neutze R, Milton DL *et al.,* (2010). Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. *PLoS Biol* **8:** e1000358.

Gómez-Pereira PR, Schüler M, Fuchs BM, Bennke C, Teeling H, Waldmann J *et al*., (2011) Genomic content of uncultured Bacteroidetes from contrasting oceanic provinces in the North Atlantic Ocean. *Environ Microbiol* **14:**52-66

Gómez-Pereira PR, Fuchs BM, Alonso C, Oliver M, van Beusekon J, Amann R. (2010). Distribution patterns and diversity of planktonic Flavobacterial clades in contrasting water masses of the North Atlantic Ocean. *ISMEJ* **4**: 472-487

Gónzalez JM, Fernández-Gómez B, Fernàndez-Guerra A, Gómez-Consarnau L, Sánchez O, Coll-Lladó M *et al.,* (2008). Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (Flavobacteria). *Proc Natl Acad Sci USA* **105**:8724-8729.

González JM, Pinhassi J, Fernández-Gómez B, Coll-Lladó M, González-Velázquez M, Puigbó P *et al.,* (2011). Genomics of the proteorhodopsin-containing marine flavobacterium *Dokdonia* sp. MED134. *Appl Environ Microbiol* **77**:8676-8686

Haft DH, Selengut J, Mongodin EF, Nelson KE. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* **1:** e60.

Hou S, Saw JH, Lee KS, Freitas TA, Belisle C, Kawarabayasi Y *et al.,* (2004). Genome sequence of the deep-sea gamma-proteobacterium *Idiomarina loihiensis* reveals amino acid fermentation as a source of carbon and energy. *Proc Natl Acad Sci USA* **101:** 18036-18041.

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D *et al.,* (2009). InterPro: the integrative protein signature database. *Nucleic Acids Res* **37:** D211-D215.

Jarrell KF, McBride MJ. (2008). The surprisingly diverse ways that prokaryotes move. *Nat Rev Micro* **6:** 466-476.

Kimura H, Young CR, Martínez A, DeLong EF. (2011). Light-induced transcriptional responses associatd with proteorhodopsin-enhanced growth i marine flavobacterium. *ISMEJ* **5:** 1641–1651

Kirchman D. (2002). The ecology of Cytophaga-Flavobacteria in aquatic environments. *FEMS Microbiol Ecol* **39:** 91-100.

Krogh A, Larsson Br, von Heijne G, Sonnhammer EL. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol* **305:** 567-580.

Langille MGI, Brinkman FSL. (2009). IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* **25:** 664-665.

Langille MGI, Hsiao WWL, Brinkman FSL. (2010). Detecting genomic islands using bioinformatics approaches. *Nat Rev Micro* **8:** 373-382.

Letunic I, Bork P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23:** 127-128.

Letunic I, Bork P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* **39**: W475-W478.

Martens EC, Koropatkin NM, Smith TJ, Gordon JI. (2009). Complex Glycan Catabolism by the Human Gut Microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem* **284:** 24673-24677.

McBride MJ, Xie G, Martens EC, Lapidus A, Henrissat B, Rhodes RG *et al.,* (2009). Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. *Appl Environ Microbiol* **75:** 6864-6875.

Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J *et al.,* (2003). GenDB--an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res* **31:** 2187-2195.

Pedrotti ML, Beauvais S, Kerros ME, Iversen K, Peters F. (2009). Bacterial colonization of transparent exopolymeric particles in mesocosms under different turbuelence intensities and nutrient conditions. *Aquat Microb Ecol* **55:** 301-312.

Pinhassi J, Sala MM, Havskum H, Peters F, Guadayol O, Malits A *et al.,* (2004). Changes in bacterioplankton composition under different phytoplankton regimens. *Appl Environ Microbiol* **70:** 6753-6766.

Pommier T, Canback B, Riemann L, Bostrom KH, Simu K, Lundberg P *et al.,* (2007). Global patterns of diversity and community structure in marine bacterioplankton. *Mol Ecol* **16:** 867-880.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies Jr *et al.,* (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35:** 7188-7196.

Pushker R, Mira A, Rodriguez-Valera F (2004). Comparative genomics of gene-family size in closely related bacteria. *Genome Biol* **5:** R27.

Qin Q-L, Zhang X-Y, Wang X-M, Liu G-M, Chen X-L, Xie B-B *et al.,* (2010). The complete genome of *Zunongwangia profunda* SM-A87 reveals its adaptation to the deep-sea environment and ecological role in sedimentary organic nitrogen degradation. *BMC Genomics* **11:** 247.

Richter M, Lombardot T, Kostadinov I, Kottmann R, Duhaime M, Peplies J *et al.,* (2008). JCoast - A biologist-centric software tool for data mining and comparison of prokaryotic (meta) genomes. *BMC Bioinformatics* **9:** 177.

Schut F, de Vries EJ, Gottschal JC, Robertson BR, Harder W, Prins RA *et al.,* (1993). Isolation of typical marine bacteria by dilution culture: growth, maintenance, and characteristics of isolates under laboratory conditions. *Appl Environ Microbiol* **59:** 2150-2160.

Stamatakis A, Hoover P, Rougemont J. (2008). A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol* **57:** 758-771.

Steindler L, Schwalbach MS, Smith DP, Chan F, Giovannoni SJ. (2011). Energy starved Candidatus *Pelagibacter ubique* substitutes light-mediated ATP production for endogenous carbon respiration. *PLoS ONE* **6:** e19725.

Stingl U, Desiderio RA, Cho J-C, Vergin KL, Giovannoni SJ. (2007). The SAR92 Clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. *Appl Environ Microbiol* **73:** 2290-2296.

Tatusov RL, Koonin EV, Lipman DJ. (1997). A Genomic perspective on protein Ffamilies. *Science* **278:** 631-637.

Woyke T, Xie G, Copeland A, González JM, Han C, Kiss H *et al.,* (2009). Assembling the marine metagenome, one cell at a time. *PLoS ONE* **4:** e5299.

Yooseph S, Nealson KH, Rusch DB, McCrow JP, Dupont CL, Kim M *et al.,* (2010). Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* **468:** 60-66.

SUPPORTING INFORMATION

Table 1SM. Enzymes and proteins identified in the four Bacteroidetes involved in nitrogen metabolism (E-value ≤ 10^{-4}).

	Locus tag			
Gene/protein	**MED152**	**MED134[a]**	**KT0803**	**MED217**
Glutamate synthase	1191-1192	c2_1337-1338	2876-2877	2043-2042
Ammonium channel (*amt*B)	1188 1190	c2_1335	487	2046
Nitrogen regulatory protein P-II	1189	- - -	488 1175 1322	2045 493
Glutamine synthase type I	-	-	-	-
Glutamine synthase type II	1212	c1_735	465	910
Glutamine synthetase type III	1211	c1_736	-	911
Glutamate dehydrogenase	1747	c2_537	3148	135
Carbon-nitrogen hydrolase	1443 1895	c1_1223 c2_179	2097 2153 3485	1811 2368 3534
Allophanate hydrolase	2366-2365	c1_545-544	2102-2103	1968-1969
Nitrous-oxide reductase (nosZ)	-	-	1398	-
Nitroreductase	232 1018 1329	c1_175 c1_883	2168 2638	1090 2761
Nitrite and Sulphite reductase	1258	-	324	446
Nitrogen regulatory protein NtrY	1900	c2_288	3177	3564
Dipeptidyl peptidase IV	1277 1345	c1_204 c2_283	2937	701 2508
POT family (Oligopeptide transporter)	1279	c1_205	2935 2936	702 703
AGCS family (Na-Ala symporter)	1007	c1_505 c1_533 c2_1460	2346 2366 2371 2633	1645 2828 2830
BCCT family:				
GlyBet	-	-	2774	843
Chol/GlyBet	-	-	2524	2550
Urea transporter and assimilation	-	-	-	-

[a]**c1 and c2 refer to the two contigs of *Dokdonia* sp. MED134's genome.**

Table 2SM. Enzymes and proteins identified in the four Bacteroidetes involved in phosphorous metabolism (E-value ≤ 10^{-4}).

	Locus tag			
Gene/protein	**MED152**	**MED134[a]**	**KT0803**	**MED217**
Alkaline Phosphatase (periplasmatic)	2526	-	2850	766
Phosphate porin	-	-	3154	2637
Phosphate permease	1915	c2_785	3153	3501
Nucleoside H^+ symporter	1066	c2_482	1012	118
Na^+ dependent nucleoside transporter	2384	c1_46	287	415
Nucleoside phosphorilase	1398	c1_181	3565	3208
Nicotinamide mononucleotide transport	1960	c2_588	3021	240 2395
PolyP/ATP NAD kinase	1458	c2_64	214	3236
Polyphosphate kinase	-	c1_975 c1_1320	1608 1817	1247 2155
PolyP glucokinase	428 1044	c1_923	2717	850
Exopolyphospatase	950	c1_1319	1816 3454	2154
H^+-translocating pyrophosphatase	2377	c1_54	294	422
Soluble pyrophosphatase	2378	c1_53	293 2732	421 1864
Phosphate permease (Phosphate starvation inducible protein, proH)	2621	c1_530	1890	1643
5'-nucleotidase (SurE) Stress condition	1285	c1_213	2929	708
Periplasmatic 5'-nucleotidase	-	c1_453 c1_454	634 635 1263 1266 1269	1616 1617 1168 1165
Periplasmatic nuclease	144 2094 2263	c1_1508 c2_171 c2_1456	2329 2527 2578 3125	2800 285 1194 1195 3541

[a]c1 and c2 refer to the two contigs of *Dokdonia* sp. MED134's genome.

Table 3SM. Enzymes and proteins identified in the four Bacteroidetes involved in sulfur metabolism (E-value ≤ 10^{-4}).

	Locus tag			
Gene/protein	**MED152**	**MED134[a]**	**KT0803**	**MED217**
Sulphatase	118	c2_1482	356 1682	286
Sulphate symporter	1965	c2_141	3553	965
Sulphate transporter	1820 2576	c2_560 c2_1490	3376 2956	99 2772
Sulphate adenyltransferase	1259-1260	c2_1489-1488	322-321	2753-2754
Adenylyl-sulfate kinase	-	-	-	2755
Phosphoadenylyl-sufate reductase	1261	c2_1487	320	445
Adenylyl-sulfate reductase	-	-	-	-
Sulfite reductase (ferredoxin)	1258	-	324	446
DMSP metabolism	-	-	-	-

[a]c1 and c2 refer to the two contigs of *Dokdonia* sp. MED134's genome.

Table 4SM. Number of Na^+ transporters identified in the four Bacteroidetes (E-value ≤ 10^{-4}).

	Number of ORFs			
Type	**MED152**	**MED134**	**KT0803**	**MED217**
ABC transporter	3	3	2	2
Na^+/H^+ exchanger	3	3	4	3
Na^+/H^+ antiporter:				
nhaA	-	-	1	1
nhaC	2	1	1	1
nhaB	-	-	-	1
Na^+/Ala symporter	1	3	4	3
Na^+/Ca^{+2} exchanger	1	1	2	2
Na^+/SO_4^{-2} symporter	1	1	1	1
Na^+:solute symporter family protein (SSF)	2	3	4	6
Na^+/bile acid cotransporter (SBF)	2	-	1	1
Na^+/HCO_3^- (SbtA)	1	-	-	-
Na^+/HCO_3^- (SulP type)	1	1	1	1
Electron chain (A-F subunits)	6	6	6	6
Na^+:dicarboxylate symporters(SDF)	1	1	1	2

Table 5SM. Predicted number of two component histidine kinases and response regulators. Number is based on hits to specific PFAMs at the Microbial Signal Transduction Database (http://genomics.ornl.gov/mist/; E-value lower than 10^{-4}). Number per megabase is shown between parentheses.

Feature	MED152	MED134	KT0803	MED217
Response regulators by themselves	25 (9)	27 (8)	37 (10)	38 (9)
His kinases by themselves	21 (7)	23 (7)	25 (7)	27 (6)
Hybrid two component systems	5 (2)	5 (2)	7 (2)	11 (3)
Total His kinases	26 (9)	28 (9)	32 (9)	38 (9)
Total	**51**	**55**	**69**	**76**

Table 6SM. Detection of signal transduction proteins (one-component system) detected in the four Bacteroidetes (E-value ≤ 10^{-4}).

	Number of ORFs			
Output domains	**MED152**	**MED134**	**KT0803**	**MED217**
HTH_AraC (AraC family protein)	11	12	12	17
GerE (LuxR family protein)	5	1	6	4
AsnC_trans_reg (AsnC/Lrp family protein)	3	5	5	5
GntR (GntR family protein)	2	1	1	1
Crp (Crp family protein)	1	1	2	3
HTH_1	0	0	1	1
HTH_3	8	8	13	15
HTH_5	1	2	3	3
HTH_8	1	1	1	1
HTH_11	1	0	0	1
LytTR	5	1	0	1
MarR (MarR family protein)	2	2	3	4
MerR (MerR family protein)	2	2	2	2
PadR	2	1	1	0
TetR_N (TetR family protein)	2	5	2	4
Trans_reg_C	0	1	0	0
HD	7	7	9	6
Guanylate_cyc	0	1	1	1
Rrf2	1	1	3	1
Input+Output domains				
LacI+Peripla_BP_1 (LacI family protein)	4	1	4	6
HTH_1+LysR LysR family protein	3	1	1	2
Crp+cNMP_binding	1	1	2	1
Fe_de_repress+Fe_dep_repr_C	1	1	1	1
HTH_11+DeoR	0	0	0	1
HTH_AraC+BLUF	1	0	0	0
PAS_3+GerE	0	0	0	1
Total	**64**	**56**	**73**	**81**

Table 7SM. Genes and domains with a potential role in adhesion (E-value ≤ 10^{-4}).

	Number of ORFs			
Domain	MED152	MED134	KT0803	MED217
Fasciclin	2	3	3	5
Thrombospondin type 3 repeat	1	2	2	1
Fibronectin typeIII	1	3	3	1
FG-GAP	4	4	3	7
Cadherin	0	0	1	0
HYR	2	1	4	0
Bac_surface_Ag	8	8	5	8
Big_1	0	0	0	0
Big_3	0	0	0	0
Dockerin_1	0	0	0	0
Cohesin	0	0	0	0
F5_F8_Type_C	3	1	1	5
He_PIG	0	0	1	0
Total	**21**	**22**	**22**	**27**

Table 8SM. The twenty most abundant paralogous families of the four Bacteroidets and two Proteobacteria (E-value ≤ 10^{-4}).

	Domain MED152	#	Domain MED134	#	Domain KT0803	#	Domain MED217	#	Domain E. *coli*	#	Domain *S. pomeroyi*	#
1	**Two component system**	33	**Two component system**	30	**Two component system**	44	**Two component system**	54	**ABC_transp**	77	**ABC_transp**	116
2	**TonB-dep_Rec**	24	**ABC_transp**	28	**TonB-dep_Rec**	39	**TonB-dep_Rec**	50	**LysR_substrate**	33	**LysR_substrate**	58
3	**ABC_transp**	22	**TonB-dep_Rec**	18	**ABC_transp**	27	**ABC_transp**	30	**Response_reg**	27	**adh_short**	30
4	adh_short	16	OmpA	13	adh_short	21	adh_short	19	EAL+GerE/MASE+GGDEF	26	Hypothetical	22
5	PKD domain	10	Peptidase_M36 (M14)	11	HlyD	19	SusD_RagB	19	Two component system	23	3HCDH_N	21
6	DEAD	8	Usp	11	SusD_RagB	13	rve	18	Fer4/ Pyr_redox_2	20	DAO/GCV_T	19
7	Sigma_70	6	PKD domain	10	hyalin/PKD domain	12	HlyD	17	Peripla_BP_1+ Lacl	18	AMP_binding	19
8	Acyl_CoA_dh_1	6	hyalin/cadherin	8	Sigma_70	12	ACR_tran	15	AA_permease	16	Two component system	17
9	Sensor protein	6	Aldedh	8	ACR_tran	11	Transposase_11	14	adh_short	16	DctM	17
10	Tranketolase	6	adh_short	8	hypothetical	10	Glyco_hydro_3	12	rve/ Pectactin	14	SBP_bac_5	17
11	GTP_EFTU	6	DEAD	8	Transposase_20	9	Glyco_hydro_43	12	Molybdoprotein	13	Acyl_CoA_dh_1	17
12	NAD-dep. aldehyde dh	6	Response_reg	7	Phage_integrase	9	Transposase_8	10	ADH_N	13	AnsC_trans_reg	16
13	PKD domain	6	Transket	7	Response_reg	9	Glyco_hydro_2_N	9	HlyD	12	DUF6	16
14	Cys_Met_Meta_PP	5	hypothetical	6	Glycos_transf_1	8	tonB-dep. receptor	9	Usher	12	Aminotran_1_2	15
15	hypothetical	5	Peptidase_M56	6	OmpA	8	rve	8	Trasposase_11	11	BPD_transp_1	15
16	SusD_RagB	5	Epimerase	6	rve	7	OmpA	8	Aldedh	11	Ald_Xan_dh_C2	14
17	hypothetical	5	GTP_EFTU	6	MS_channel	7	Transposase_20	8	DeoR+ HTH_DeoR	11	Aldedh	13
18	HTH_AraC	5	Aminotran_1_2	5	DEAD	7	Transposase_8	8	BPD_transp_1	10	BPD_transp_1	11
19	SpoU_methylase	5	Sigma_70	5	E1_E2_ATPase	7	E7_E2_ATPase	7	BPD_transp_1	10	BPD_transp_1	11
20	PseudoU_synth_2	5	Aminotran_3	5	Pyr_redox_2	7	Response_reg	7	PTS_EIIC	10	HTH_AraC	10

Table 9SM. Clusters of genes putatively involved in the attachment and degradation of polymeric compounds for the four Bacteroidetes. Only those that contain the tandem TonB dependent/ligand-gated channel genes and SusD are represented. Notice the amount of genes involved in degradation, adhesion, carbohydrate metabolism, secretion and transport within the clusters. Cells with a minus sign have no Pfam assigned at E-value ≤ 10^{-4}. Empty cells mean no information available.

ORF MED152	Annotation	Pfam domain	Direction	SigP	TMHMM
109	Alpha-amylase	Alpha-amylase	⇐	yes	0
110	HP	-	⇐	yes	0
111	HP	-	⇐	no	0
112	Outer membrane protein	**SusD_RagD**	⇐	yes	0
113	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	0
114	Transcriptional regulator, LacI family	Peripla_BP_1	⇐	no	0
115	Major Facilitator Superfamily	MSF_1	⇒	no	11
116	ß-phosphoglucomutase	Hydrolase	⇒		0
117	Trehalase/maltose phosphorylase	Glyco_hydro_65m	⇒		0
118	N-acetylglycosamine-6-sulfatase	Sulfatase	⇒	no	0
119	Alpha-amylase	Alpha-amylase	⇒		0
120	Putative esterase	Esterase	⇒	yes	0
121	Glycosyl hydrolase, family 31	Glyco_hydro_31	⇒	yes	0
122	Alpha-amylase	Alpha-amylase	⇒	no	0
123	Alpha-amylase	Alpha-amylase	⇒	yes	0
124	Conserved HP	Y_Y_Y	⇒	yes	1
125	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	0
126	Outer membrane protein	**SusD_RagD**	⇒	yes	0
127	Conserved HP	-	⇒	no	0
128	HP	-	⇒	yes	0
129	Conserved HP	-	⇒	yes	0
130	Conserved HP	-	⇒		0
131	Conserved HP	-	⇒	no	0
132	Glycosyl hydrolase family 30	Glyco_hydro_30	⇒	yes	0
133	Major Facilitator Superfamily	MSF_1	⇒	no	12

1050	Major Facilitator Superfamily	MSF_1	⇐	no	12
1051	Trehalase/maltose hydrolase	Glyco_hydro_65m	⇐	yes	0
1052	HP	Cellulase	⇐	yes	0
1053	Alpha-trehalase	Trehalase	⇐		0
1054	Sugar transporter	MSF_1	⇐	no	12
1055	Sucrose transporter	MSF_1	⇐	no	12
1056	Conserved HP	-	⇐	no	0
1057	Conserved HP	UnbV_ASPIC	⇐	yes	0
1058	Conserved HP	UnbV_ASPIC	⇐	no	0
1059	HP	-	⇐	yes	1
1060	Conserved HP	UnbV_ASPIC	⇐	no	0
1061	Outer membrane protein	**SusD_RagD**	⇐	yes	0
1062	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	1
1063	Conserved HP	DUF1080	⇐	yes	0
1064	Conserved HP	AP_endonuc_2	⇐	no	0
1065	Oxidoreductase	GFO_IDH_MocA	⇐	no	0
1066	HP	Nuc_H_symport	⇐	no	12
1067	Conserved HP	AP_endonuc_2	⇐	yes	0
1068	Transcriptional regulator, AraC	HTH_AraC	⇐		0
1963	S-adenosylhomocysteine hydrolase	AdoHcyase	⇒		0
1964	Putative sulfite reductase	-	⇒	yes	4
1965	Putative sodium/sulphate symporter	Na_sulph_symp	⇒	no	13
1966	Putative alginate lyase precursor	-	⇒	no	0
1967	Putative alginate lyase precursor	-	⇒	yes	0
1968	Gluconate lyase, SKI family	SKI	⇒		0
1969	6-phosphogluconate dehydrogenase	6PGD	⇒	no	0
1970	Putative Mn+2 Fe+2 transporter	Nramp	⇒	yes	11
1971	Putative alginate lyase precursor	-	⇒	no	0
1972	HP	adh_short	⇒	no	0
1973	HP	-	⇒	yes	0
1974	HP	Hepar_II_III	⇒	no	0
1975	HP	Cupin_2	⇒		0

1976	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	0
1977	Outer membrane protein	**SusD_RagD**	⇒	no	0
1978	HP	-	⇒	no	0
1979	HP	PKD	⇒	yes	1
1980	Transcriptional regulator, GntR	FCD	⇒		0
1981	Major Facilitator Superfamily protein	MSF_1	⇒	no	11
1982	Short-chain dehydrogenase/ reductase	adh_short	⇒	no	0
1983	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	0
1984	Outer membrane protein	**SusD_RagD**	⇒	yes	0
2482	Putative tRNA/rRNA methyltransferase	SpoU_methylase	⇐		0
2483	HP	-	⇐	no	0
2484	Outer membrane protein	**SusD_RagD**	⇐	yes	0
2485	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	1
2486	HP	-	⇐	no	0
ORF MED134	**Annotation**	**Pfam domain**	**Direction**	**SigP**	**TMHMM**
c1_1083	HP	-	⇐		
c1_1084	HP	Peptidase_M56	⇐	no	4
c1_1085	Transcriptional regulator	Penicillinase_R	⇐		0
c1_1086	Putative hydrolase	Peptidase_M20	⇐	yes	0
c1_1087	Outer membrane protein	**SusD_RagD**	⇐	no	0
c1_1088	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	0
c1_1089	Putative hydrolase	Peptidase_M20	⇐	yes	0
c1_1090	Probable aminopeptidase	Peptidase_M28	⇐	yes	0
c1_1091	HP	DUF1083	⇐	yes	0
c1_1092	HP	-	⇐	yes	1
c1_1093	HP	-	⇐	yes	0
c1_1094	HP	-	⇐	no	2
c1_1095	UvrABC system protein A	ABC_tran	⇐	no	0
c2_32	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	1
c2_33	Outer membrane protein	**SusD_RagD**	⇒	yes	0

c2_34	Putative multidrug resistance protein	ACR_tran	⇐	no	14
c2_35	Putative HlyD-like secretion protein	HlyD	⇐	yes	0
c2_36	Putative outer membrane efflux protein	OEP	⇐	yes	0
c2_37	Transcriptional regulator, TetR	TetR_N	⇐		0
c2_668	HP	-	⇐	yes	0
c2_668	HP	-	⇐	no	0
c2_670	HP	MS_channel	⇐	no	5
c2_671	HP	Peptidase_M56	⇐	yes	3
c2_672	Putative antibiotic resistance-related regulatory protein	Penicillinase_R	⇐		0
c2_673	Putative succinyl-diaminopimelate desuccinylase	Peptidase_M20	⇐		0
c2_674	2-keto-3-deoxy-6-phosphogluconate aldolase	Aldolase	⇐		0
c2_675	2-dehydro-3-deoxygluconokinase	Pfk	⇐	no	0
c2_676	Short-chain dehydrogenase/ reductase	adh_short	⇐	yes	0
c2_677	Putative hexuronate transport protein	MSF_1	⇐	no	11
c2_678	Transcriptional regulator GntR family	FCD	⇐		0
c2_679	Putative lyase	-	⇐	yes	0
c2_680	Cell surface protein	PKD	⇐	yes	0
c2_681	Outer membrane protein	**SusD_RagD**	⇐	yes	0
c2_682	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	1
c2_683	Putative pectin degradation protein	Cupin_2	⇐		0
c2_684	Putative chondroitin AC/ alginate lyase	-	⇐	no	0
c2_685	Putative chondroitin AC/ alginate lyase	-	⇐	yes	0
c2_686	Putative chondroitin AC/ alginate lyase	-	⇐	yes	0
c2_687	HP	-	⇐	no	0

c2_696	Putative two-component system sensor (hydrid)	His_kinase	⇒	no	4
c2_697	Response regulator receiver protein	Response_reg	⇒		0
c2_698	HP	-	⇒	yes	0
c2_699	Putative carbohydrate kinase	PfkB	⇐		0
c2_700	Xylose transporter	Sugar_tr	⇐	no	12
c2_701	Glycoside hydrolase, family 32	Glyco_hydro_32N	⇐	no	0
c2_702	Transcriptonal regulator, LacI	Peripla_BP_1	⇒	no	0
c2_703	Beta-phosphoglucomutase	Hydrolase	⇒		0
c2_704	Trehalase/maltose hydrolase	Glyco_hydro_65m	⇒		0
c2_705	HP	-	⇐	no	0
c2_706	HP	Pentapeptide	⇐		0
c2_707	HP	AraC_E_bind	⇐	no	1
c2_708	HP	-	⇐	yes	0
c2_709	ASPIC-like protein	UnbV_ASPIC	⇐	no	0
c2_710	ASPIC-like protein	UnbV_ASPIC	⇐	no	0
c2_711	ASPIC-like protein	UnbV_ASPIC	⇐	yes	0
c2_712	Outer membrane protein	**SusD_RagD**	⇐	yes	0
c2_713	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	0
ORF KT0803	**Annotation**	**Pfam domain**	**Direction**	**SigP**	**TMHMM**
9	Major Facilitator Superfamily permease	Sugar_tr	⇒	yes	12
10	Glycosyl hydrolase, family 32	Glyco_hydro_32N	⇒	yes	0
11	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	0
12	Outer membrane protein	**SusD_RagB**	⇒	yes	0
13	Carbohydrate kinase	PfkB	⇒	no	0
14	Multidrug resistance protein	Multi_Drug_Res	⇒	no	0
15	Metal-dependent membrane protease	Abi	⇒	no	8
16	Hyaluronan synthase	Glycos_transf_2	⇐		5
17	HP	-	⇐		0
18	Membrane protein	-	⇒	no	2
19	Secreted protein	Guanylate_cyc	⇐	yes	1
20	Secreted protein	-	⇐	yes	0
21	Major Facilitator Superfamily permease	MSF_1	⇒	no	11

22	Secreted protein	-	⇒	yes	1
23	Conserved HP	-	⇐	yes	1
24	Secreted protein	-	⇐	yes	0
25	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase	E1_E2_ATPase	⇐		8
26	Secreted protein	Cation_efflux	⇐	no	5
27	Conserved HP	-	⇐		0
28	Fur family transcriptional regulator protein	-	⇐		0
29	HlyD family secretion protein	-	⇐	no	0
30	Heavy metal cation efflux protein	ACR_tran	⇐	no	13
31	HP	-	⇐	no	1
32	Beta-galatosidase	Glyco_hydro_2_N	⇐	yes	0
33	Arabinogalactan 1,4-beta-galactosidase	Glyco_hydro_53	⇐	yes	0
34	Secreted protein	-	⇐	yes	0
35	Outer membrane protein	**SusD_RagB**	⇐	no	0
36	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	1
37	Two-component system sensor (hydrid)	HATPase_c	⇐	yes	1
38	Membrane protein	-	⇒	no	4
39	Short-chain dehydrogenase/reductase	adh_short	⇒	no	0
357	Sensor/regulator hydrid	GerE	⇒	yes	1
358	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	0
359	Outer membrane protein	**SusD_RagB**	⇒	no	0
360	Conserved HP	-	⇒	yes	0
361	Conserved HP	-	⇒	no	0
362	Secreted alpha/beta fold hydrolase	Abhydrolase_2	⇒	yes	1
363	Fibronectin III containing domain GH43	F5_F8_type_C	⇐	yes	0
364	Glycosyl hydrolase family 43	Glyco_hydro_3	⇒	yes	1
365	Glycosyl hydrolase family 2	Glyco_hydro_2_N	⇐	yes	0
366	Secreted protein	-	⇐	yes	0
367	Secreted lipase/esterase	-	⇐	yes	0

678	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	1
679	Outer membrane protein	**SusD_RagB**	⇒	no	0
680	Alpha-L-arabinofuranosidase	Alpha_L_AF_C	⇒	yes	0
681	Arabian endo-1,5-L-arabinosidase	Glyco_hydro_43	⇒	yes	0
682	Conserved HP, membrane or secreted	DUF1680	⇒	yes	1
683	Glycosyl hydrolase, family 43	Glyco_hydro_43	⇒	no	0
684	Arabian endo-1,5-L-arabinosidase	Glyco_hydro_43	⇒	yes	0
685	Alpha-L-arabinofuranosidase	Glyco_hydro_43	⇒	yes	0
686	HP	-	⇒		0
687	Ribulokinase	FGGY_C	⇒		0
688	L-ribulose-5-phsopate 4-epimerase	Aldolase_II	⇒	no	0
689	L-arabinose isomerase	Arabinose_Isome	⇒	no	0
690	Aldose 1-epimerase	Aldose_epim	⇒	no	0
691	Sodium:solute symporter family protein	SSF	⇒	no	14
692	Transaldolase	Transaldolase	⇒	no	0
693	Transketolase N-terminal	Transketolase_N	⇒		0
694	Transketolase C-terminal	Trasnket_pyr	⇒		0
695	Alpha-glucosidase	-	⇒	yes	0
696	Protein containing tetratricopeptide repeats	TRP_2	⇐	no	0
697	HP	-	⇐		1
698	Outer membrane efflux protein	OEP	⇐	no	0
699	ABC-type transporter ATP-binding protein	ABC_tran	⇐	no	0
700	Predicted permease	FtsX	⇐	yes	8
701	Conserved HP	-	⇐	yes	0
702	Predicted permease	FtsX	⇐	no	8
703	Predicted permease	FtsX	⇐	yes	8
704	Predicted permease	FtsX	⇐	no	8
705	Predicted permease	FtsX	⇐	yes	8
706	ABC-type transporter ATP-binding protein	ABC_tran	⇐		0
707	HlyD family secretion protein	-	⇐	yes	1

708	Two-component system response regulator	Sigma54_activat	⇒	no	0
709	Two-component system sensor histidine kinase	HATPase_c	⇒	yes	2
1014	Glucoside 3-dehydrogenase	-	⇒	no	0
1015	Conserved HP	-	⇒	no	0
1016	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	0
1017	Outer membrane protein	**SusD_RagB**	⇒	no	0
1018	Secreted protein	UnbV_ASPIC	⇒	yes	0
1019	Secreted protein	UnbV_ASPIC	⇒	no	1
1020	HP	-	⇒	no	0
1021	HP	AraC_E_bind	⇒	no	1
1135	Secreted alginate lyase-like protein	-	⇒	yes	0
1136	HP	-	⇒	no	0
1137	Secreted protein	-	⇒	yes	0
1138	Secreted protein	Hepar_II_III	⇒		0
1139	Pectin degradation protein	Cupin_2	⇒		0
1140	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	0
1141	Outer membrane protein	**SusD_RagB**	⇒	yes	0
1142	Protein containing PDK domain	PKD	⇒	no	0
1143	Secreted alginate lyase-like protein	-	⇒	yes	0
1144	Secreted alginate lyase-like protein	-	⇒	yes	0
1145	Trancriptional regulator, GntR	FCD	⇒		0
1146	Major Facilitator Superfamily permease	MSF_1	⇒	no	11
1147	Short-chain dehydrogenase/ reductase	adh_short	⇒	no	0
1148	2-dehydro-3-deoxygluconokinase	PfkB	⇒		0
1149	Aldolase	Aldolase	⇒	no	0
1150	Fructose-1,6-bisphosphatse	FBPase	⇒	no	0
1151	Transcriptional regulator, LacI	Peripla_BP_1	⇒	no	0
1152	Phage integrase protein	Phage_integrase	⇒	no	0
1153	Secreted protein	-	⇒	yes	0

1154	Membrane protein	DUF305	⇒	no	3
1241	FerR family protein	FecR	⇒	no	1
1242	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	1
1243	Outer membrane protein	**SusD_RagB**	⇒	yes	0
1244	Transmembrane spermine/ spermidine synthase	Spermine_synth	⇒	no	13
1245	HP	-	⇒	no	0
1246	HP	-	⇒		0
1247	RNA polymerase ECF-type sigma factor	Sigma70_r2	⇒		0
1248	FerR family protein	FecR	⇒		1
1249	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	no	0
1250	Outer membrane protein	**SusD_RagB**	⇒	yes	0
1251	Conserved HP, secreted	-	⇒	yes	0
1252	Secreted protein	-	⇒	no	2
1253	Conserved HP, membrane	DUF81	⇒	no	8
1367	Peptidase family 16	Peptidase_M16_C	⇐	no	0
1368	Outer membrane protein	**SusD_RagB**	⇐	no	0
1369	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	1
1686	L-fucose permease	MSF_1	⇐	no	12
1687	Short-chain dehydrogenase/ reductase	adh_short	⇐	no	0
1688	Amidohydrolase family protein	Aminohydro_2	⇐		0
1689	Altronate oxidoreductase	Mannitol_dh_C	⇐		0
1690	Altronate hydrolase	GD_AH_C	⇐		0
1691	Aldolase	Aldolase	⇐		0
1692	2-dehydro-3-deoxyluconokinase	PfkB	⇐		0
1693	Zinc-type alcohol dehydrogenase	ADH_N	⇐		0
1694	Alpha-L-fucosidase	Alpha_L_fucos	⇐	yes	0
1695	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	no	0
1696	Outer membrane protein	**SusD_RagB**	⇒	no	0
1697	Protein containing PDK domains	PKD	⇒	no	1

1698	Secreted glycosyl hydrolase, family 16	Glyco_hydro_16	⇒	yes	0
1699	Membrane or secreted glycosyl hydrolase	CMB_4_9	⇒	yes	0
1700	Phytanoyl-CoA dioxygenase family protein	PhyH	⇒		0
1701	Zinc-type alcohol dehydrogenase	ADH_N	⇒		0
1702	Acetylesterase	Esterase	⇒		0
1703	Secreted glycosyl hydrolase, family 2	Glyco_hydro_2_N	⇒	yes	1
1704	Secreted protein	-	⇒	yes	0
1705	Secreted protein	DUF1080	⇒	yes	0
1706	Nucleotidyltransferase family protein	NTP_transf_2	⇒		0
1707	HP	-	⇒		0
1708	Conserved HP	-	⇒	no	0
1709	Glycosyl hydrolase, family 16	CMB_4_9	⇐	no	0
1710	HP	-	⇐	no	0
1711	Membrane or secreted glycosyl hydrolase	Glyco_hydro_2_N	⇐	yes	0
1712	Secreted glycosyl hydrolase, family 16	Glyco_hydro_16	⇐	yes	0
1713	HP	-	⇒		0
1714	Secreted beta-lactamase family protein	Beta-lactamase	⇒	yes	0
1715	Conserved HP	-	⇐		0
1716	Outer membrane protein	**SusD_RagB**	⇐	no	0
1717	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	0
1718	FecR family protein	FecR	⇐		1
1719	RNA polymerase ECF-type sigma factor	Sigma70_r4_2	⇐		1
1720	Periplasmatic trehalase-like protein	Trehalase	⇒	yes	0
1721	Aminoglycoside phospho trasferase	Aminotran_3	⇒		0
1722	Beta-galactosidase	Glyco_hydro_2_C	⇒	yes	0
2111	Glyceraldehyde-3-phosphate dehydrogenase A	Gp_dh_C	⇐	no	0
2112	6-phosphofructokinase	PFK	⇐	no	0
2113	Alpha amylase	Alpha_amylase	⇐	no	0
2114	Alpha amylase	Alpha_amylase	⇐	yes	0

2115	Glycoside hydrolase, family 65	Glyco_hydro_65m	⇐		0
2116	Beta-phosphoglucomutase	Hydrolase	⇒	no	0
2117	Major Facilitator Superfamily permease	MSF_1	⇒	no	12
2118	Transcriptional regulator, LacI	Peripla_BP_1	⇒	yes	0
2119	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	0
2120	Outer membrane protein	**SusD_RagB**	⇒	yes	0
2121	Conserved HP	-	⇒	yes	0
2122	Alpha amylase	Alpha_amylase	⇒	yes	0
2123	Ribonucleotide reductase large subunit	Ribonuc_red_IgC	⇒		0
2814	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	0
2815	Outer membrane protein	**SusD_RagB**	⇒	yes	0
2816	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	4
2817	Conserved HP	-	⇒	yes	0
2818	Secreted protein	-	⇒	yes	0
2819	HP	-	⇒	no	1
2820	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	1
2821	Outer membrane protein	**SusD_RagB**	⇒	yes	0
2883	HP	-	⇐		0
2884	Conserved HP	-	⇐	yes	0
2885	HP	-	⇐	yes	0
2886	Outer membrane protein	**SusD_RagB**	⇐	no	0
2887	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	1
3418	Two-component system senso (hydrid)	Y_Y_Y	⇒	yes	1
3419	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	1
3420	Outer membrane protein	**SusD_RagB**	⇒	yes	1
3421	Conserved HP	-	⇒	no	0
3422	Glycosyl hydrolase, family 16	Glyco_hydro_16	⇒	yes	1
3423	Beta-glucosidase	Glyco_hydro_3_C	⇒	yes	0

3424	Glycosyl hydrolase, family 16	Glyco_hydro_16	⇒	yes	0
3533	Gliding motility protein	DUF21	⇒	no	3
3534	Gliding motility protein	-	⇒	yes	0
3535	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	no	0
3536	Outer membrane protein	**SusD_RagB**	⇒	yes	0
3537	Membrane protein	DUF6	⇐	no	10
3538	Secreted peptidase, family M16	Peptidase_M16_C	⇐	yes	0
3539	Secreted peptidase, family M16	Peptidase_M16_C	⇐	yes	0
ORF MED217	**Annotation**	**Pfam domain**	**Direction**	**SigP**	**TMHMM**
284	Dipeptidyl peptidase 4	Peptidase_S9	⇐		1
285	Endonuclease	Exo_endo_phos	⇐		0
286	N-acetylglucosamien-6-sulfatase	Sulfatase	⇐		0
287	Alpha/beta hydrolase fold	-	⇐	yes	0
288	Glycosyl hydrolase, family 3	Glyco_hydro_3	⇐	yes	0
289	Carboxyesterase	Abhydrolase_2	⇐		0
290	HP	-	⇐		0
291	HP	-	⇐		0
292	Outer membrane protein	**SusD_RagB**	⇐		0
293	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	0
344	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒		0
345	Outer membrane protein	**SusD_RagB**	⇒		0
346	Alkaline phosphatase	Alk_phosphatase	⇒	yes	0
347	Alkaline phosphatase	Alk_phosphatase	⇒		0
348	Endo-1,4-beta-xylanase	Glyco_hydro_43	⇐	yes	0
349	TonB dependent receptor	**TonB_dep_Reg**	⇒	yes	0
350	HP	-	⇒		0
351	Two-component sensor histidine kinase	HATPase_c	⇐		2
352	Transcriptional regulator	Sigma54_activat	⇐		0
353	HP	-	⇒		1
354	ABC transporter, APT-binding protein	ABC_tran	⇒		0
355	Predicted permease	FtsX	⇒		8

356	Predicted permease	FtsX	⇒	yes	8
357	Predicted permease	FtsX	⇒		8
358	Predicted permease	FtsX	⇒	yes	8
359	Predicted permease	FtsX	⇒		8
360	Predicted permease	FtsX	⇒	yes	8
361	ABC transporter, AT-binding protein	ABC_tran	⇒		0
362	Outer membrane efflux protein	OEP	⇒		0
363	Pirin family protein	Pirin	⇒		0
364	Bacterial regulatory protein	GerE	⇒	no	0
365	HP	Peptidase_S41	⇒		0
366	Two-component sensor histidine kinase	HATPase_c	⇒		1
614	ASPIC-like protein	UnbV_ASPIC	⇐		0
615	ASPIC-like protein	UnbV_ASPIC	⇐		0
616	ASPIC-like protein	UnbV_ASPIC	⇐		0
617	Outer membrane protein	**SusD_RagB**	⇐	no	0
618	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	0
683	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒		0
684	Outer membrane protein	**SusD_RagB**	⇒		0
685	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	no	0
686	Outer membrane protein	**SusD_RagB**	⇒	no	0
687	HP	-	⇒	0	0
922	HP	DUF718	⇐		0
923	HP	-	⇐		0
924	Sulfatase	Sulfatase	⇐		0
925	Outer membrane protein	**SusD_RagB**	⇐		0
926	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐		0
927	Transforming growth factor	Fasciclin	⇐		0
928	Transforming growth factor	Fasciclin	⇐		0
929	Outer membrane protein	**SusD_RagB**	⇐	yes	1
930	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	no	0
931	Transforming growth factor	Fasciclin	⇐		

1069	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒		0
1070	Outer membrane protein	**SusD_RagB**	⇒	yes	0
1071	HP	-	⇒		0
1072	Glycosyl hydrolase, family 16	Glyco_hydro_16	⇒		0
1073	HP	-	⇒		0
1074	HP	-	⇒		0
1075	Beta-glucosidase	Glyco_hydro_3C	⇒		0
1076	Glycosyl hydrolase, family 16	Glyco_hydro_16	⇒	yes	0
1193	Endonuclease	Exo_endo_phos	⇐		0
1194	Endonuclease	Exo_endo_phos	⇐		0
1195	Outer membrane protein	**SusD_RagB**	⇐		0
1196	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	0
1197	Two-component system sensor histidine kinase	HATPase_c	⇐		1
1198	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒		0
1199	Outer membrane protein	**SusD_RagB**	⇒		0
1200	HP	-	⇒		0
1201	HP	-	⇒		0
1202	Xylosidase	Glyco_hydro_43	⇒	no	0
1327	Alpha/beta hydrolase fold	Abhydrolase_1	⇒		8
1328	Bacterial regulatory family LuxR protein	GerE	⇒		0
1329	Probable TonB-dependent outer membrane receptor	-	⇒		0
1330	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	no	1
1331	Outer membrane protein	**SusD_RagB**	⇒		0
1332	Carboxyesterase	COesterase	⇒		0
1333	Alpha-L-rhamnosidase	Bac_rhamnosid	⇒		0
1334	Two-component system response regulator	Response_reg	⇒		0
1335	Two-component system sensor histidine kinase	His_kinase	⇐	no	2
1336	Phosphatase	Metallophos	⇐		1

1337	Beta-glucuronidase	Glyco_hydro_2N	⇒		1
1338	Hydrolase	Lactamase_B	⇐		0
1543	Outer membrane protein	**SusD_RagB**	⇐		0
1544	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐		0
1545	Putative beta-xylosidase	Glyco_hydro_43	⇐		0
1546	Two-component system sensor histidine kinase	HATPase_c	⇒		0
1547	HP	Glyco_hydro_43	⇐		0
1548	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐		1
1549	Outer membrane protein	**SusD_RagB**	⇐		0
1550	HP	-	⇐		0
1551	HP	-	⇐	yes	0
1552	HP	-	⇐		0
1553	HP	-	⇐		0
1554	Beta-galactosidase	Glyco_hydro_2C	⇐		0
1555	Alpha-galactosidase	-	⇐	yes	0
1556	Putative secreted hydrolase	Glyco_hydro_16	⇐		0
1557	Putative glycosyl hydrolase	Glyco_hydro_88	⇐	yes	0
1558	HP	-	⇐		0
1559	Beta-galactosidase	Glyco_hydro_2C	⇐		0
1560	HP	-	⇐		1
1561	Aldo/keto reductase family protein	Aldo_ket_red	⇐	no	0
1562	Endo-1,4-beta-xylanase A	-	⇒	no	0
1563	Putative membrane protein	MSF_1	⇒		12
1572	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒		0
1573	Outer membrane protein	**SusD_RagB**	⇒	no	0
1998	HP	-	⇐	no	0
1999	HP	-	⇐	yes	0
2000	Outer membrane protein	**SusD_RagB**	⇐		0
2001	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐		0
2002	HP	-	⇐		0
2003	Two-component system (hybrid)	HATPase_c	⇐	no	1

2318	Aldolase 1-epimerase	Aldolase_epim	⇒		0
2319	L-arabinose isomerase	Arabinose_isome	⇐		0
2320	Sodium/glucose contransporter 1	SSF	⇐	no	14
2321	L-ribulose-5-phosphate 4-epimerase	Aldolase_II	⇐		0
2322	L-ribulokinase	FGGY_C	⇐		0
2323	Alpha-N-arabinofuranosidase A	Alpha_L_AF_C	⇐		0
2324	HP	DUF1680	⇐		0
2325	Arabinan endo-1,5-alpha-L-arabinosidase A	Glyco_hydro_43	⇐		0
2326	Xylosidase/arabinosidase	Glyco_hydro_44	⇐	yes	0
2327	Arabinan endo-1,5-alpha-L-arabinosidase A	Glyco_hydro_43	⇐		0
2328	Alpha-L-arabinosidase	Alpha_L_AF_C	⇐		0
2329	ASPIC-like protein	UnbV_ASPIC	⇐		0
2330	Outer membrane protein	**SusD_RagB**	⇐	no	0
2331	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐		1
2587	Mannonate dehydratase	UxuA	⇐		0
2588	Short-chain dehydrogenase/ reductase	adh_short	⇐		0
2589	Outer membrane protein	**SusD_RagB**	⇐		0
2590	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	0
2591	Trascriptional regulator, LacI	Peripla_BP_1	⇐		0
2592	Hypotethical oxidoreductase	GFO_IDH_MocA	⇐		0
2593	HP	Aminohydro_2	⇐		0
2594	Two-component system sensor histidine kinase	HATPase_c	⇒	no	2
2595	Probable transcriptional regulator	Response_reg	⇒		0
2596	Xylosidase/arabinosidase	Glyco_hydro_43	⇒		0
2665	Glucokinase	ROK	⇐		1
2666	Esterase	Abhydrolase_3	⇐		0
2667	Putative alpha-glucosidase	-	⇐		0
2668	HP	Glyco_hydro_92	⇐	yes	0
2669	HP	DUF1237	⇐		0

2670	HP	-	⇐		0
2671	HP	F5_F8_type_C	⇐	no	0
2672	Outer membrane protein	**SusD_RagB**	⇐		0
2673	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐		0
(…)	HP	-	⇐		
2681	Putative alpha-1,2-mannosidase	Glyco_hydro_92	⇐		0
2682	L-fucose-proton symporter	MSF_1	⇐		12
2683	Putative esterase	Esterase	⇐		1
2684	Arylsufatase	Sulfatase	⇐	no	0
2685	Beta-galactosidase	Glyco_hydro_2	⇐		0
2686	Beta-galactosidase	Glyco_hydro_2_N	⇐	0	
2687	Two-component system (hybrid)	HATPase_c	⇒	yes	1
2688	Rhamnogalacturonan acetylesterase precursor	Lipase_GDSL	⇒	yes	0
2689	Beta-galactosidase	Glyco_hydro_2_N	⇐		1
2690	HP	Lipase_GDSL	⇐		0
2691	Acetylesterase	DUF303	⇐		0
2692	Arabinose metabolism trascriptional repressor	GntR	⇐		0
2693	Hypothetical oxidoreductase	adh_short	⇒		0
2694	Putative sugar isomerase	-	⇒		0
2695	Glycerol kinase	FGGY_N	⇒		0
2696	L-lactate dehydrogenase	FMN_dh	⇒		0
2697	Transmembrane protein	-	⇒		11
2698	Exo-poly-alpha-D-galacturonidase precurson	Glyco_hydro_28	⇒		0
2699	HP	-	⇒		1
2700	HP	-	⇒	yes	0
2701	HP	-	⇒		0
2702	HP	-	⇒		0
2703	Arabinosidase	-	⇐		0
2704	Xylanase	Peptidase_S9	⇐	yes	1
2705	Beta-galactosidase	Glyco_hydro_2_N	⇐		0
2706	Glycosyl hydrolase, family 88	Glyco_hydro_88	⇐		0
2707	Conserved HP	Glyco_hydro_2_N	⇐	yes	0
2708	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒		0

2709	Outer membrane protein	**SusD_RagB**	⇒		0
2710	Two-component system sensor histidine kinase	HATPase_c	⇒		0
2711	HP	-	⇒		0
2712	HP	-	⇒		0
2713	HP	-	⇒	no	0
2714	HP	Response_reg	⇐		0
2715	Two-component system sensor histidine kinase	HATPase_c	⇐		0
2716	HP	-	⇐		0
2717	Aminotransferase	Aminotran_1_2	⇐		0
2718	HP	-	⇒		0
2719	Transcriptional regulator, AraC	HTH_AraC	⇒		0
2720	HP	-	⇒		5
2721	HP	-	⇒		1
2722	HP	Peptidase_S9	⇐		0
2723	Outer membrane protein	**SusD_RagB**	⇐	no	0
2724	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐		1
3416	Ribonucleoside-diphosphate reductase alpha chain	Ribonuc_red_IgC	⇐		0
3417	Ribonucleoside-diphosphate reductase beta chain	ATP-cone	⇐	no	0
3418	Alpha-glucosidase	Glyco_hydro_31	⇐		0
3419	Alpha-amylase	Alpha_amylase	⇐	yes	0
3420	Alpha-amylase	Alpha_amylase	⇐	yes	0
3421	HP	-	⇐		0
3422	Outer membrane protein	**SusD_RagB**	⇐		0
3423	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	no	1
3424	Serine/threonine-protein kinase	-	⇐		1

Chapter 3

Patterns and architecture of genomic islands in marine bacteria

Beatriz Fernández-Gómez, Antoni Fernàndez-Guerra, Emilio O. Casamayor, José M. González, Carlos Pedrós-Alió, and Silvia G. Acinas

ABSTRACT

Genomic Islands (GIs) have an important role in modulating the bacterial genome structure and size displaying a diverse set of laterally transferred genes. However, GIs in marine bacterial genomes have not been explored systematically to uncover possible trends and to analyze their putatively ecological significance. In this study, we performed a comprehensive analysis of GIs in 70 selected marine bacterial genomes to explore the distribution, patterns and functional gene content in these genomic regions. We detected 438 GIs containing a total of 8152 genes. We showed that the GI number per genome was strongly and positively correlated with the total GI size. In 50% of the genomes analyzed the GIs accounted for approximately 3% of the genome length, with a maximum of 12%. Interestingly, we found transposases particularly enriched within Alphaproteobacteria GIs, and site-specific recombinases in Gammaproteobacteria GIs. We described specific Homologous Recombination GIs (HR-GIs) in several genera of marine Bacteroidetes and in *Shewanella* strains among others. In these HR-GIs we recurrently found housekeeping genes such as β-subunit of DNA-directed RNA polymerase, regulatory sigma factors, the elongation factor Tu, and ribosomal protein genes typically associated with the core genome. Our results indicate that both horizontal gene transfer mediated by phages, plasmids and other mobile genetic elements, and HR by site-specific recombinases or plasmids, play important roles in the mobility of clusters of genes between taxa and within closely related genomes, modulating the flexible pool of the genome. Our findings suggest that GIs may increase bacterial fitness under environmental changing conditions by acquiring novel foreign genes and/or modulating gene transcription and/or transduction.

INTRODUCTION

Bacterial comparative genomics is providing a unique opportunity to retrieve valuable information regarding genome structure, functional diversity and evolution of marine microorganisms. Bacterial genomes are dynamic entities with a known pool of conserved housekeeping genes at the core genome that remain similar and comparable at different taxonomic levels, and the flexible (or adaptive) genome, with a number of genes that are not comparable among closely related strains (Hacker and Carniel 2001; Ochman *et al*., 2005). Horizontal gene transfer (HGT) is one of the evolutionary mechanisms enlarging the flexible genome pool of bacterial populations, facilitating their adaptation to new ecological niches (Doolittle 1999; Boucher *et al.,* 2003). The flexible genome pool has been analyzed for a few bacterial taxa by comparative genome analysis (Hacker and Carniel 2001). In well-known marine bacteria such as *Prochlorococcus*, *Synechococcus*, and *Shewanella* comparative genome analysis has revealed a substantial number of species-specific genes (Kettler *et al.,* 2007; Dufresne *et al*., 2008; Konstantinidis *et al*., 2009). Those studies revealed an unsaturated pangenome size, reflecting the existence of new lineages and the heterogeneity among the flexible genome pool. Species-specific genes are commonly found in genomic islands (GIs), that are defined as laterally transferred clusters of genes linked to the flexible pool of the genome (Hacker and Kaper 2000; Hacker and Carniel 2001).

GIs are important genomic regions causing significant genetic differences between closely related genomes, and they may reveal particular ecologically relevant features of the genomes (Cuadros-Orellana *et al*., 2007), and virus-bacteria interactions (Coleman *et al*., 2006; Avrani *et al*., 2011). GIs may harbor a large set of genes with different origins. However, using a hypothesis-free approach for identification of GIs some common features can be recognized suggesting that GIs could be perceived as a superfamily of mobile elements (Vernikos and Parkhill 2008). Genes found within GIs are very diverse: from key genes for survival in specific environments to virulence, and/or antibiotic resistance genes. In fact, GIs enriched in virulence genes and in Clustered Regularly Interspaced Palindromic Repeats (CRISPRs) confer resistance to exogenous genetic elements such as plasmids and phages (Ho Sui *et al.,* 2009). Thus, GI content may hold clues about the lifestyle or survival strategies of bacteria (Read and Ussery 2006). Another common characteristic of the GIs is the enrichment in novel genes without any orthologous groups detected in the COG database or any other known functional gene families (Hsiao *et al*., 2005).

Extensive literature exists on GIs in pathogenic bacterial strains (referred to as pathogenic islands) where their relevance is known for antibiotic resistance or virulence

stages (Hacker and Carniel 2001; Schmidt and Hensel 2004; Gal-Mor and Finlay 2006). In environmental microorganisms GIs have been associated with the presence of catabolic pathways for organic pollutants, thus conferring adaptive traits in some *Pseudomonas* sp. strains (van der Meer and Sentchilo 2003). Other ecological features associated with GIs are the presence of genes for magnetite biomineralization in what is called the magnetosome island in the Alphaproteobacterium *Magnetospirillum gryphiswaldense* (Ullrich *et al*., 2005), or secondary metabolism in marine Actinobacteria strains (Penn *et al.,* 2009). Another case is the acquisition of a capsular polysaccharide biosynthesis gene cluster by the non-pathogenic soil inhabitant *Burkholderia thailandensis* with similar characteristics to the virulence gene cluster of the pathogenic *Burkholderia pseudomallei* (responsible for the melioidosis disease) (Sim *et al.,* 2010). In Cyanobacteria GIs from several strains of *Prochlorococcus* marinus (Coleman *et al*., 2006; Kettler *et al*., 2007) and *Synechococcu*s sp. strains (Dufresne *et al*., 2008) have been reported. Also, the GIs of two freshwater filamentous toxin-producing cyanobacteria were found with diverse comparative approaches (Stucken *et al.,* 2010). For Gammaproteobacteria GIs were described for the high pressure adapted *Photobacterium profundum* SS9 strain (Campanaro *et al.,* 2005), the marine coastal *Vibrio vulnificus* (Cohen *et al*., 2007), *Alteromonas macleodii* (Ivars-Martínez *et al.,* 2008) and *Shewanella baltica* strains (Caro-Quintero *et al.,* 2011). In Alphaproteobacteria GIs were found in SAR11 (Candidatus *Pelagibacter ubiqu*e strain HTCC1062) referred to as hypervariable regions (HVR) in the original study (Wilhelm *et al*., 2007). In aquatic Bacteroidetes GIs or HVR were described in *Salinibacter ruber*, a highly abundant bacterium in solar salterns, (Pasic *et al.,* 2009; Peña *et al*., 2010). Finally, virulence genes of typically pathogenic island were reported in marine bacteria genomes in a comparative study (Persson *et al.,* 2009). However, GIs in marine bacterial genomes have not been explored systematically and a comparative analysis is still lacking.

In this study we performed a comprehensive analysis of GIs in 70 selected marine bacterial genomes that represented abundant and ecologically relevant bacteria in the ocean. We assembled a database of 8152 genes found in GIs of marine bacteria and screened it for possible patterns and clues about the ecological relevance of GIs in marine bacteria.

MATERIALS AND METHODS

Genomic island prediction and database construction

Seventy prokaryotic genomes were analyzed in this study. GIs of 53 genomes were obtained at the time of our analyses (February 2010) directly from IslandViewer database (http://wwwpathogenomics.sfu.ca/islandviewer). IslandViewer is a web-based interface that integrates several methods for identification and visualization of genomic islands: IslandPick, IslandPath-DIMOB and SIGI-HMM (Langille and Brinkman 2009). IslandPick (Langille *et al.,* 2008) is a comparative genomic GIs prediction method that requires phylogenetically related genomes to be available for the comparison. SIGI-HMM measures codon usage (Waack *et al*., 2006) and IslandPath (Hsiao *et al.,* 2003) the abnormal sequence composition or the presence of genes related to mobile elements to identify possible GIs. For a recent review of bioinformatics approaches to detect GIs see Langille *et al.* (2010). The remaining 17 genomes (eight Alphaproteobacteria and nine marine Bacteroidetes) were downloaded from the National Center for Biotechnology Information (NCBI, http://ncbi.nlm.igh.gov) and the J. Craig Venter Institute (https://moore.jcvi.org/moore). Fourteen of these 17 genomes were not closed. The genomes that were not available in IslandViewer were uploaded and the GIs predicted by IslandPick by selecting closely related genomes (at least 3 genomes when possible) plus one reference distant genome. IslandPick GIs prediction detects also the GIs that overlap with the other two GIs predictors, IslandPath-DIMOB and SIGI-HMM (Langille and Brinkman 2009) and this information was also integrated in our database. IslandViewer detects GIs ≥8 kb although only those ≥9.5 kb were used to compile our database for further analyses to be consistent with previous published studies. GIs of the 70 bacterial genomes were exported in csv format and all proteins in a fasta file. The genome selection criteria used were: (i) presence of marine genomes with pre-calculated GIs within IslandViewer, (ii) availability when possible of at least three closely related genomes, and (ii) widest taxonomic representation of ecologically relevant marine bacteria.

Phylogenetic Analyses

Maximum likelihood (ML) phylogenetic analyses of selected genomes were carried out on full-length 16S rRNA gene sequences, and the elongation factor Tu (EF-Tu) and the ß subunit of the RNA polymerase (*rpo*B) on amino acid sequences. The sequences were aligned using MAFFT [v.6.857] with the algorithm E-INS-I (Katoh and Toh 2008). An additional more stringent alignment was constructed by removing ambiguously aligned sites using Gblocks (Castresana 2000) as well as visual examination. Phylogenies were constructed using ML as

implemented in RAxML [v.7.2.8] (Stamatakis *et al.,* 2008) with GTR nucleotide substitution model for the 16S rRNA sequences and the BLOSUM62 amino acid substitution matrix for the EF-Tu and rpoB sequences. The trees generated were visualized and edited in Interactive Tree Of Life (Letunic and Bork 2007).

Control (manually) vs. automatically detected genomic islands analyses

We selected eight genomes found in the public sequence databases as control genomes, in which the GIs had been manually annotated, to evaluate the accuracy of our approach for GI prediction. We calculated several parameters such as accuracy (% overlap), precision, recall (sensitivity) as detailed in Langille *et al.* (2008) shown in Table 1. The GIs of control genomes were verified by different approaches (see Table 1). Comparison of the control and automatically predicted genomic islands was carried out using the genomic display software CIRCOS (Krzywinski *et al.,* 2009).

Functional Annotation of Genomic Islands

Gene sequences found in the predicted genomic islands were extracted and stored in a flat file-type database. The potential protein domains were analyzed using the package HMMER 3.0 (Eddy 1998) against the PFAM 24.0 database (Finn *et al.,* 2010). Clusters of orthologous groups of proteins (COGs) for characterization of the proteins (Tatusov *et al.,* 1997) were carried out using the rpsblast program bundled in the NCBI BLAST package using an e-value of $1e^{-5}$ as threshold (Altschul *et al.,* 1990). In addition, Blast2GO (B2G) was used to run the functional annotation of the extracted sequences with an e-value of $1e^{-20}$ and cut-off identity in their amino acid sequences of 55% (Conesa *et al.,* 2005; Gotz *et al.,* 2008). Gene ontologies (Ashburner *et al.,* 2000) and EC number from KEGG pathways (Kanehisa and Goto 2000) were retrieved to identify the main biological processes, molecular functions, and cellular components present in the genomic islands.

Statistical and GIs comparison analyses

Both total genome peptides and peptides within genomic islands were classified by BLASTP hits to the NCBI's Cluster of Orthologous Genes (COG) as described in Lauro *et al.* (2009) with cutoffs of E value $\leq 1e^{-5}$, identity ≥ 30 and coverage ≥ 50%. Hypothetical proteins were considered as those with no hits to the COG database. To estimate protein functions that were over-represented in different taxonomic groups we used the Gossip package (Blüthgen *et al*., 2005) implemented in the bioinformatics annotation tool Blast2GO

(Conesa *et al.,* 2005). This package uses the Fisher´s Exact Test with multiple testing corrections and, since multiple categories are examined simultaneously, the Benjamini and Hochberg False Discovery Rate correction (FDR) for multiple testing was determined for all functional category analyses. We considered p-values smaller than 0.01 to be significant ($p < 0.01$). Fisher´s Exact Test (FT) was used to explore putative significant differences at three comparative levels: (i) the functional gene annotation of the combined GIs of each control genome and their corresponding automatically detected GIs, (ii) the combined GIs dataset for the eight control genomes and their corresponding automatically detected GIs and (iii) the comparative gene enrichment analyses among the bacterial taxa of Cyanobacteria, Gammaproteobacteria, Alphaproteobacteria and marine and non marine Bacteroidetes to detect functional categories enriched within each taxa.

In addition, we assigned genes within GIs to 16 biological categories defined by us based on the most representative GO terms with ecological relevance and related to: photosynthesis (1) (photosystem, antenna proteins and electron transport system), energy metabolism enzymes (2), ribosomal proteins (3), hydrolysis (4), polysaccharide biosynthesis (5), DNA restriction modification system (Type I, II & III) (6), DNA-directed RNA polymerases (7), transporters (ABC and multridrug/metal resistence) (8), two component system (9), cell motility (flagellum and chemotaxis) (10), stress response (heat shock or chaperone proteins) (11), MGE (Conjugative transposon, integrases, phage integrases, transposon Tn21/Tn7) (12); CRISPRs (13), virulence (14), TA toxins (plasmid killer system) (15), and secretion systems proteins (Type I, II and III) (16). GO terms enrichment analyses and the matrix of ecological gene categories, were implemented in the iTOL software for visualization (http://itol.embl.de).

RESULTS AND DISCUSSION

Accuracy of GI prediction: Manually (control) vs. automatically detected genomic islands

It is impractical and excessively time consuming to manually curate 70 genomes in order to identify all the GIs in them. Thus, we had to resort to an automatic search. IslandViewer has been shown to be one of the best tools for this purpose (Langille *et al.,* 2010). IslandViewer is a web-based interface that integrates several methods for identification and visualization of genomic islands: IslandPick, IslandPath-DIMOB and SIGI-HMM (Langille and Brinkman 2009). IslandPick (Langille *et al.,* 2008) is a GI prediction tool based on comparisons with

phylogenetically related genomes. The existence of well-annotated genomes that are close phylogenetic relatives of the examined genome is usually not a problem for pathogenic bacteria. However, for marine bacteria there are much less genomes sequenced and even less have been manually curated. We tried to have at least three closely related genomes for each marine genome examined, but this was not always possible. Therefore, we were limited in the accuracy of our GI calling by this lack of closely related genomes.

The other tools in IslandViewer are SIGI-HMM that relies on codon usage, based on the frequency and preferences for synonymous codons used to encode proteins in each species (Waack *et al.,* 2006) and IslandPath-DIMOB that uses the di (MGE) nucleotide bias and the occurrence of at least a single mobile genetic elements gene to identify possible GIs (Hsiao *et al.,* 2003). Again, these features are much better understood in pathogenic bacteria than in marine bacteria. Therefore, we decided to test how the GIs detected by IslandViewer compared to those present in a few genomes of marine bacteria that had been manually annotated previously. This comparison would place whatever patterns were found in the right perspective. The genomes used as controls are shown in Figure 1 and Table 1.

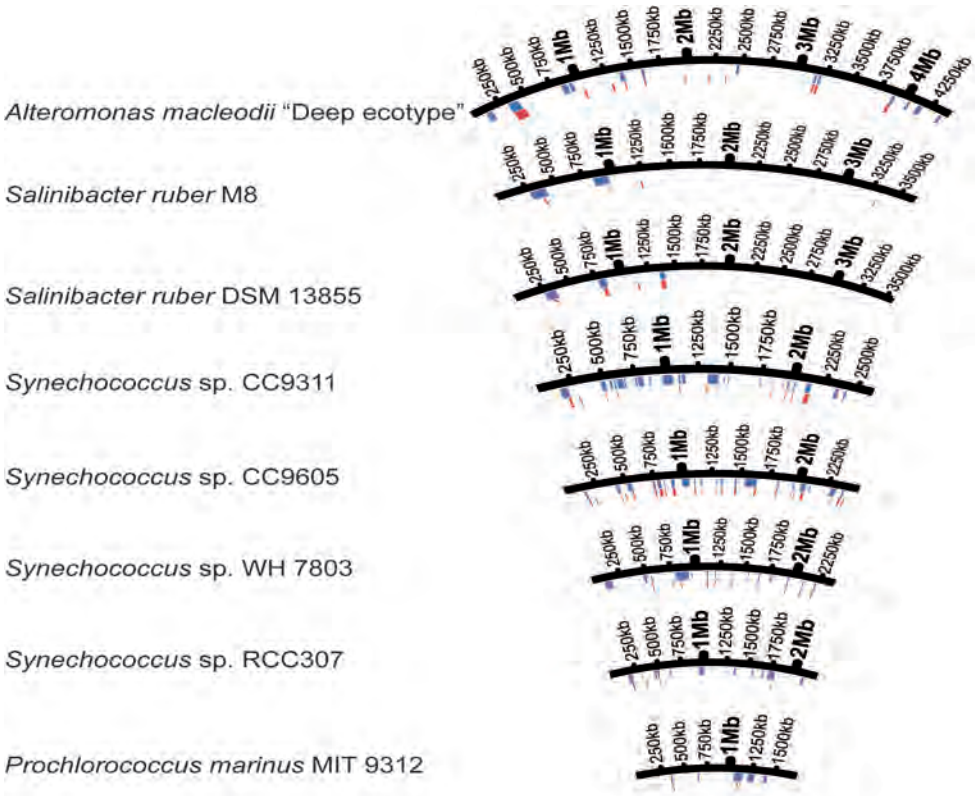


Figure 1. **Positions of the GIs in eight selected marine bacterial genomes used as controls**. Blue bars show the GIs detected manually and red bars show the GIs predicted by IslandViewer. This graphic includes all GIs detected by IslandViewer although only GIs > 9.5 kb were included in our dataset for further analyses.

The manually detected GIs are shown in blue in Figure 1 together with the automatically detected GIs in red. Obviously, IslandViewer did not detect all the GIs in the control genomes. The percentage of GIs detected ranged between 27% in *Synechococcus* sp. RCC307 and 100% in both *Salinibacter* strains. Another discrepancy was that some areas were detected as GIs that were not annotated as such in the genomes. In most cases there were only one or two "false positives" per genome, although there were five in *Alteromonas* and four in *Synechoccocus* sp. CC9605. In all cases, however, these false positives were very short (Fig. 1), usually smaller than 8 kb and therefore these GIs were not included in our final dataset. In terms of the length of DNA in GIs (in kb) IslandViewer detected approximately between 20 and 60% of the length in manually annotated GIs with an average of 36% (Table 1).

Table 1. Comparison of the GIs of 8 marine bacteria referred to as Control Genomes where GIs were manually detected and the GIs predicted automatically for the same genomes by IslandViewer/IslandPick.

	Manual		Automatic				Manual vs. Automatic GIs		
Bacterial strain	Number of GIs	Total GIs length detected in control genomes (kb)	True GIs[a] (+)	False GIs[b](+)	False GIs[c] (-)	Total GIs length with automatic prediction (kb)*	Accuracy[a] (% overlap) (kb)	Recall[b] (sensitivity) (%)	Precision[c] (%)
A. macleodii "Deep ecotype"[3]	13	480	6	5	7	272.7	69.6	46	55
S. ruber M8[4]	2	265.9	2	3	0	144.9	44.5	100	40
S. ruber DSM13855[4]	3	221.8	3	1	0	168.2	41.1	100	75
S. sp. CC9311[2]	24	578.6	9	2	15	125.9	16.9	38	82
S. sp. CC9605[2]	20	505.3	18	4	2	300.4	54.3	90	82
S. sp. WH7803[2]	11	344	3	2	8	73.7	20.3	27	60
S. sp. RCC307[2]	15	271	9	1	6	56.9	18.7	60	90
P. marinus str. MIT9312[1]	5	233	3	0	3	44.9	19.3	50	100
Average							**35.58**	**64**	**73**

[1]Coleman *et al.*, 2006. Method used to detect GIs: Comparative genomics using 2 *Prochlorococcus marinus* strains genomes.
[2]Dufresne *et al.*, 2008. Method used to detect GIs in *Synechococcus* genomes: Modified protocols published by Hsiao (Hsiao *et al.*, 2005) and Rusch et al. (Rusch *et al.*, 2007); specifically deviations of tetranucleotide frequency, presence of MGE, tRNAs and the occurrence of core genome gene blocks.
[3]Ivars-Martínez *et al.*, 2008. Method used to detect GIs: Comparative genomics using 2 *Alteromonas macleodii* strains genomes.
[4]Peña *et al.*, 2010. Method used to detect GIs: Comparative genomics using 2 *Salinibacter ruber* strain genomes.
a) Accuracy (% overlap) between manual and automatic annotations.
b) Recall (sensitivity): True +/ (True + plus False -).
c) Precision: True +/(True + plus False +).

* All GIs detected by IslandViewer/IslandPick included some small genomic islets (<9.5 kb) that are not included in Table 3SM.

These percentages were smaller than those found in a previous test based on 118 bacterial genomes (mostly pathogenic bacteria with many closely related genomes available) also using Island Viewer (88% on average) (Langille *et al.,* 2008; Langille *et al.,* 2010). However, both precision (average 73%) and sensitivity (average 64%) were very good when compared with the different methods explored in Langille *et al*. (2008) (see their Table 1). That is, the automatic detection we obtained was as good as can be with current bioinformatics tools.

In summary, when looking at our data set we will be missing a substantial portion of the GIs but, most of the sequence will belong to true GIs. This caveat will apply to any studies of bacteria that are not pathogenic and authors using Island Viewer should be aware of this difficulty.

So far, we have been concerned only with the number and length of the GIs detected by IslandViewer. Next we will consider the identity of the genes found in the GIs that were detected by both systems. We did two types of comparisons. In the first one, we considered each genome separately and, in the second one, we considered all the control genomes pooled together. Obviously, for this purpose, only genes with clearly assigned GO categories could be considered. Thus, out of all the detected genes, the hypothetical proteins were discarded for this comparison (Table 1SM). The number of genes ranged between 51 and 225 in the manually annotated subset and between 16 and 116 in the automatically annotated subset. Fischer's exact tests revealed no significant differences in the proportion of genes in each GO category between the two data subsets. This shows that the subset of GI genes retrieved by IslandViewer is a non-biased representation of all the GI genes in each control genome.

Finally, we pooled all the manually annotated genes in GIs of the eight control genomes together (1065 genes), and all the automatically predicted genes in the GIs for the same genomes (397 genes). If the different GO categories were similarly represented in both data sets, our conclusions about functional aspects of the genes in GIs would be justified. If there were significant differences, however, our conclusions would be biased by the particular subset of GIs detected by IslandViewer. We only found three specific GO terms with significantly different distributions in both datasets. One GO term related to photosynthesis (GO:0015979) was found under-represented in the automatically annotated database. This was probably due to the smaller percentage of GIs detected in some of the cyanobacterial strains, such as *Synechococcus* sp. CC9311 and RCC307 or *Prochlorococcus marinus* MIT9312. Thus, even though differences were not significant when considering each genome separately, pooling together five cyanobacterial genomes made the difference significant (Table 1).

Two other GO terms were overrepresented in the automatically annotated data set. These were related to DNA recombination (GO:0006310) and DNA binding (GO:003677). This was likely due to the fact that IslandViewer detects many GIs based on the existence of mobile genes. Likely, this was also the cause of the "false positives," since isolated mobile elements tend to be ignored during manual annotation unless they are an important objective

for the researcher (and usually they are not). Thus, IslandViewer should be more efficient at detecting these GIs than manual annotation.

In conclusion, IslandViewer detects only a subset of the islands annotated manually. Reasons for this are the different criteria used by different authors to define GIs. In many cases the GIs detection is based only on one approach; either comparison of two genomes or differences of the tetranucleotide frequency and occurrence of mobility genes (Table 1). IslandViewer seeks several characteristics and, therefore, we can expect a more robust and more conservative detection of GIs. The important conclusion for the present work, however, is that the functional representation of genes is not significantly different between the manual and automated procedures. Therefore, our conclusions will be representative of all GIs. This conclusion is important for other authors seeking to use IslandViewer for GI detection in other environmental bacteria.

Quantitative importance of GIs in marine bacterial genomes

The 70 selected marine bacterial genomes represent the four major prokaryotic taxa in the ocean: Cyanobacteria (16 genomes), Gammaproteobacteria (17), Alphaproteobacteria (16) and Bacteroidetes (21) (Table 2SM). These four bacterial taxa account for up to 80% of total marine bacterioplankton (Barberán and Casamayor 2010). Bacteroidetes genomes included 14 Flavobacteria and 7 non-marine Bacteroidetes (*Bacteroides* spp.) used as out-groups. Several genomes of closely related bacterial strains from each phylogenetic group were included to investigate the rate of variability of GIs at intra-specific level and explore their relevance as main contributors for strain-specific genes.

GIs were detected in 66 out of the 70 bacterial genomes (Table 2SM). No GIs were detected in the genomes of *Pelagibacter ubique* HTCC1062 and the marine Bacteroidetes *Flavobacteria bacterium* BBFL7, *Flavobacteriales bacterium* ALC-1, and *Flavobacterium psychrophilum* JIP02/86. The absence of GIs in these genomes may be related to the small genome size of some of them (between 1.4 and 3.8 Mb), as well as to the lack of sensitivity of the GI predictor when suitable genomes were not available for comparison. Overall, we detected 438 GIs spanning a total of 8.87 Mb (Table 2SM). The size of individual GIs ranged between 10 and 436 kb (389 kb per genome on average, considering the marine genomes only). As expected, the total GI size was strongly and positively correlated with the number of GIs per genome ($R=0.93$, $p <0.001$) (Fig. 2A). Significant correlations ($p <0.001$) were observed between genome size and both GI size and number of GIs per genome (Fig. 2B and

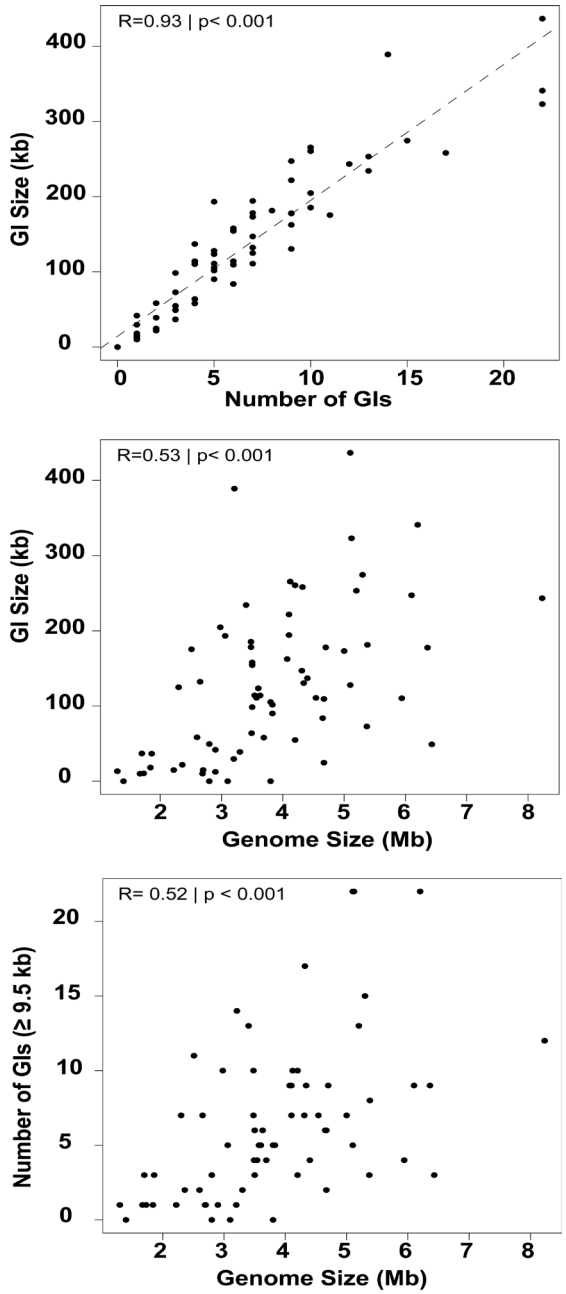


Figure 2. **Patterns of GIs in marine bacterial genomes. A)** Relationship between number of GIs per bacterial genome with GIs size. **B)** Relationship between bacterial genome and GIs size. **C)** Relationship between genome size and number of GIs (≥9 kb).

2C). When the data were analyzed separately for each of the four classes, Cyanobacteria (R= 0.69, $p < 0.05$) and Bacteroidetes (R=0.78, $p < 0.001$) showed significant correlations (Figure 1A and 1D SM) while the Proteobacteria did not (Fig. 1B and 1C SM).

For any given range of genome sizes, there was a large variability in the length of GIs (data not shown). The fraction of the bacterial genome represented by GIs ranged between 0 and 12% (Fig. 3). Most genomes showed a ratio between 2 and 5%. The Cyanobacteria showed a significantly lower average ratio than the other three classes. However, this is likely due to the low GI detection rate of Island Viewer in the case of several Cyanobacteria (Table 1). Otherwise,

there were no significant differences among classes, although the Gammaproteobacteria showed a lower variability than Bacteroidetes and Alphaproteobcteria (Fig. 3).

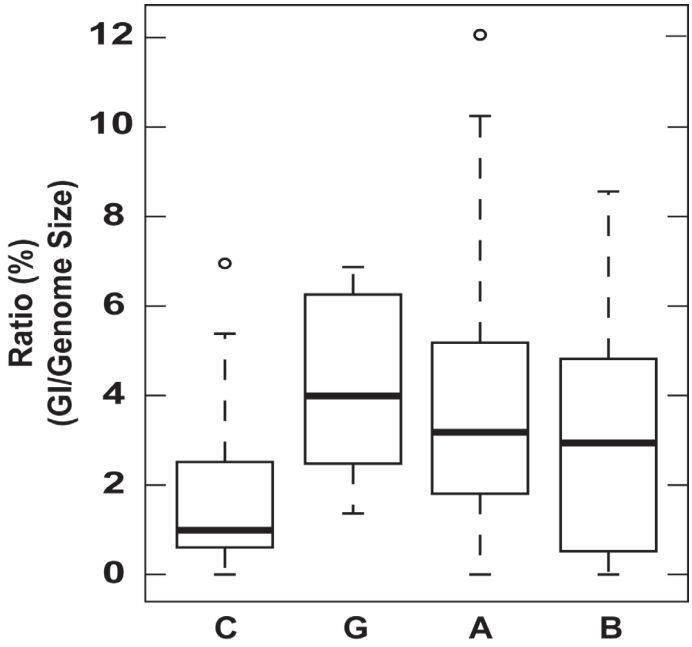


Figure 3. Box- and whiskers graphic of the GI ratio (in %) for the 70 marine bacteria. Genomes are ranked from highest (12%) to lowest (0) and grouped in 4 main phylogenetic affiliation represented as follow: **C** (Cyanobacteria), **G** (Gammaproteobacteria), **A** (Alphaproteobacteria) and **B** (Bacteroidetes). The graph shows the median (thick horizontal line), the upper and lower quartile (rectangle), the maximum and minimum values excluding outliers (discontinuous line), and finally circle represents an outlier.

The genomes with the highest ratios for each main bacterial class were the alphaproteobacterium *Rhodobacter sphaeroide*s ATCC1705 (12%), the cyanobacterium *Synechococcus* sp. CC9605 (7%), the gammaproteobacterium *Psychrobacter* sp. PRwf-1 (6.8%), and the flavobacterium *Robiginitalea biformata* HTCC2501 (4.5%; Table 2SM). In a previous study, *Synechococcus* GIs were shown to form between 10 and 31% of their genomes (Dufresne *et al.,* 2008) and similar percentages, and up to 17%, have been found also for pathogenic islands in *Escherichia coli* (Ochman *et al.,* 2000). The marine bacteria examinated have a lower percentage of their genome in GIs.

Interestingly, high intra-specific variability in GIs size was observed in some members of every group. Two *Synechococcus* sp. strains, for example, with genomes of 2.2 and 2.6 Mb respectively, showed very different GI ratios (15 kb and 175 kb respectively). Similarly, small differences in genome size between two S*hewanella baltica* strains (MR-4 and OS155, with 4.7 and 5.12 Mb respectively) contrasted with marked differences in their GIs ratios (3.7 and

6.3% respectively) (Table 2SM and Fig. 1SM).

Architecture of marine bacterial GIs

As a common characteristic we found that 70% of the described GIs contained mobile genetic elements (MGE), mostly transposases, conjugative transposons, integrons or phage integrase-related genes in accordance with previous GIs revisions (Dobrindt *et al.,* 2004, Juhas *et al.,* 2008). In addition, we observed that at least 27% of the GIs were flanked by or contained tRNAs, probably acting as the integration sites for GIs (Reiter *et al.,* 1989; Williams 2002). Most GIs in marine bacterial genomes could be assigned to one of two architectures. First, many of these GIs exhibited a high content of hypothetical proteins (HP), as well as different sets of genes in different genomes, suggesting that they had originated through

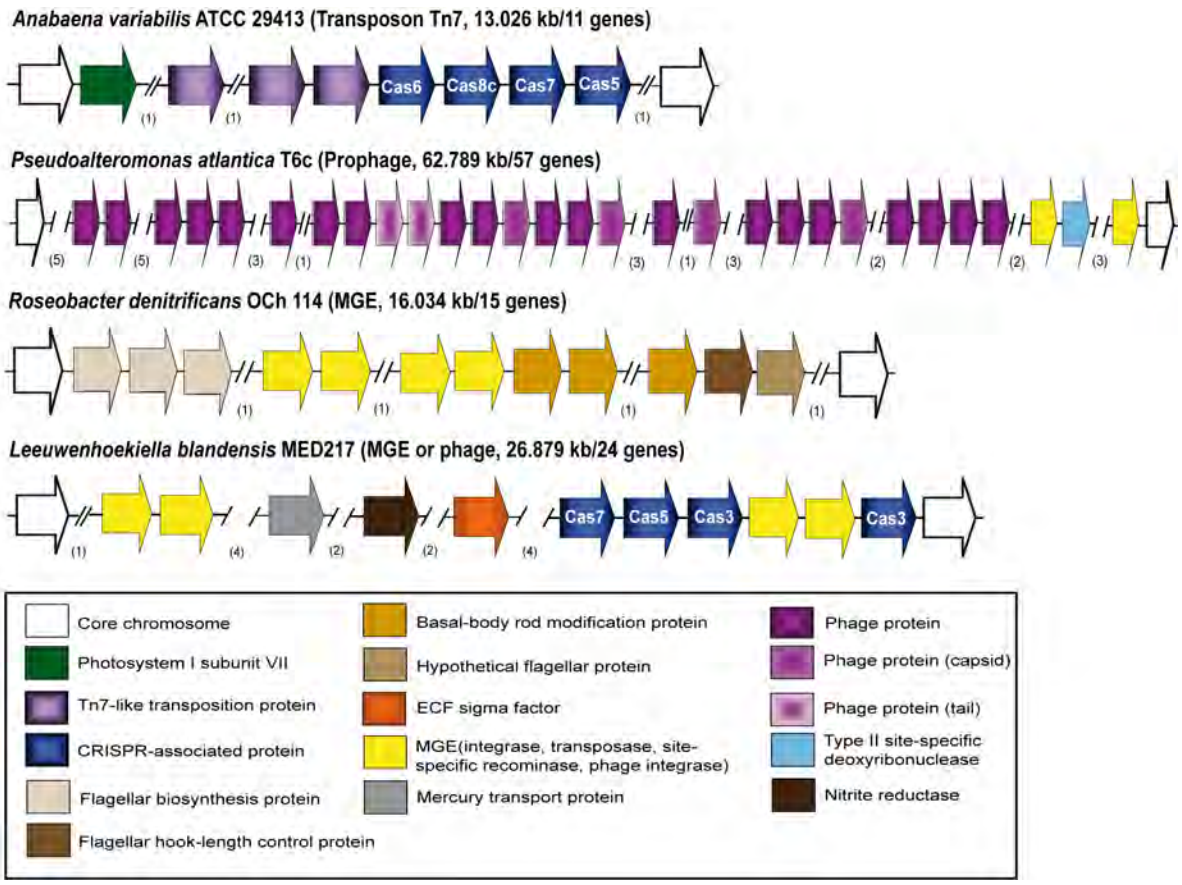


Figure 4. Structure (5´-3´) of the Horizontal Gene Transfer (HGT)-GIs representative of four marine bacteria. Genomes belonged to Cyanobacteria, Gammaproteobacteria, Alphaproteobacteria and marine Bacteroidetes. The hypothetical origin of HGT (via prophage, transposon or other MGE), the length of the GI (in kb) and the number of genes integrated are shown in brackets. Colors indicate the variety of genes observed within the GIs. Numbers in brackets under the genes indicate the HPs and other genes not considered for the figure.

horizontal gene transfer (HGT) by phages, conjugative transposons or other MGEs. From now on, we will refer to these GIs as HGT-GIs. One example from each bacterial class examined is shown in Figure 4. The cyanobacterium *Anabaena variabilis* ATCC 29413 displayed a gene related to Photosystem I, four cas genes related to CRISPRs system and three Tn7 transposition proteins in a single GI of 13 kb. The gammaproteobacterium *Pseudoalteromonas atlantica* T6c presented a GI of 62.7 kb mostly constituted by a prophage with many phage related proteins genes. Also, in the alphaproteobacteria *Roseobacter denitrificans* Och 114 we detected a GI of 16 kb with many flagellar proteins and MGE elements. Finally, the marine bacteroidetes *Leeuwenhoekiella blandensis* MED217 had a GI of 26.8 kb with multiple genes of MGE, cas genes, and a nitrite reductase. This gene has been only described in marine bacteroidetes in the deep sea flavobacterium *Zunongwangia profunda* SM-A87 with capacity to hydrolize organic nitrogen (Qin *et al.,* 2010). Many other GIs presented ecologically interesting genes but specific details for each one are out of scope of this paper although a summary of 16 biological categories is shown in Figure 10.

And second, we found many GIs that contained site-specific recombinases and tRNAs. Interestingly, these GIs were characterized by harboring many housekeeping (or core) genes, almost no hypothetical proteins, and had a structure that could be repeatedly detected in genomes of different genera or even different classes (Fig. 5). We hypothesized that such genomic fragments were likely transferred via homologous recombination (HR). We will refer to these as HR-GIs. Quite likely these genomic cassettes may be also mobilized within the same genome by the transposases that some of them have at their flanks. Flavobacteria was one of the classes with more conspicuous HR-GIs. Figure 5 shows an example of one the HR-GI named as HR1-GI. This GI was found in many Bacteroidetes (Fig. 2SM) of which five were specifically detected in our dataset and are shown in Figure 5 as an example. Despite differences in the total length of the island (from 15.9 to 44 kb), synteny was maintained for a cassette consisting of a substantial number of genes (see black rectangle in Fig. 5). It is notorious that the genes observed upstream of the cassette were quite different in every genome (Fig. 5). Surprisingly, housekeeping genes as important as the β-subunit of DNA-directed RNA polymerase, the elongation factor Tu, sigma factors, transcription termination factors, and ribosomal proteins were recurrently detected in this HR-GIs, and they also contained site-specific recombinases and tRNAs. The HR1-GI detected in five marine Flavobacteria (Fig. 5) was further examined in other Bacteroidetes representatives. We observed identical synteny in nine Flavobacteria, three Sphingobacteria and three Bacteroides (Fig. 2SM). The only variants were a few gene insertions in *Capnocythophaga ochracea* DSM 7271 and some deletions in sphingobacterial genomes. This particular HR1-GI, therefore, was well conserved throughout the Bacteroidetes phylum.

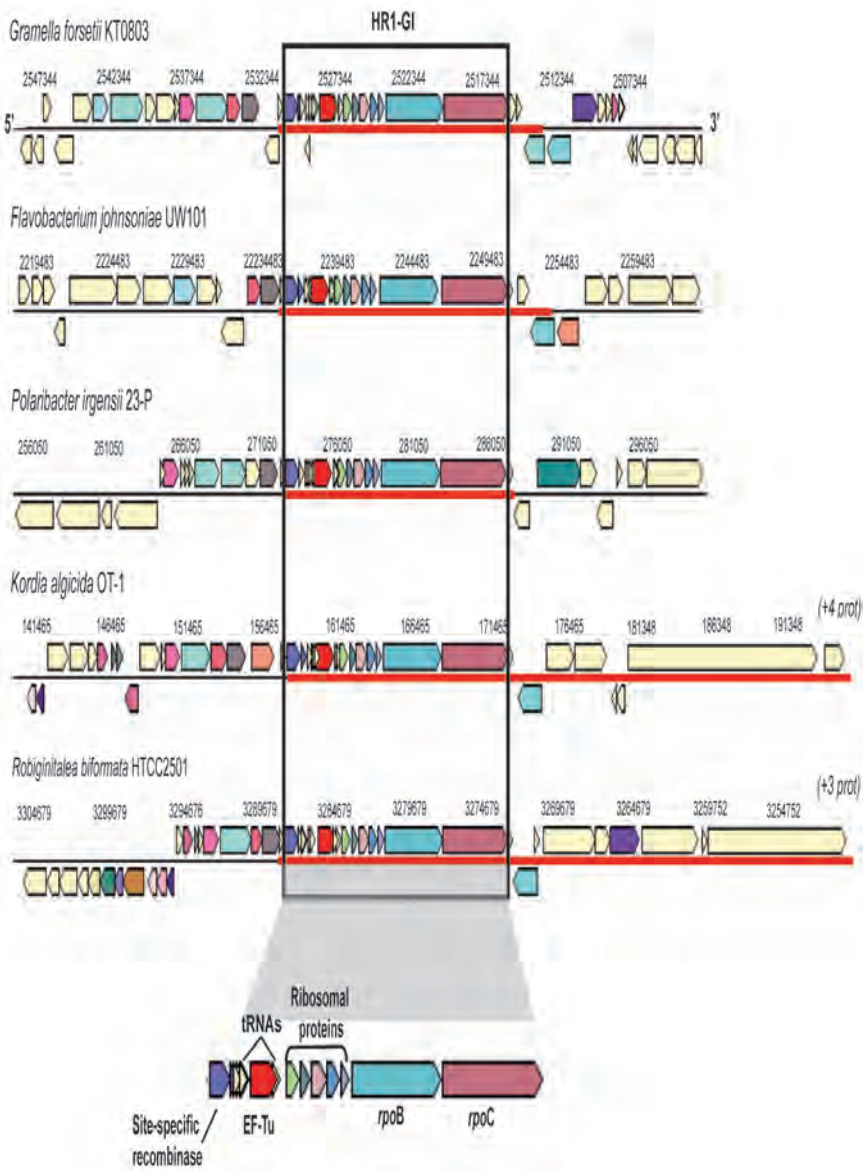


Figure 5. Structure (5´-3´) of the Homologous Recombination GIs (HR-GIs) in marine bacterial genomes. HR1-GI detected in five different genera of marine Bacteroidetes. The synteny and the gene cassette shared by these genomes are within the black box. Red line indicates the length of the GI detected which varied among the genomes.

Some of the genes present in these particular HR-GIs coded for ribosomal proteins, which are known to be highly expressed genes, usually with a sequence composition very different from the rest of the genome (Karlin 2001). As a consequence, these genome fragments might appear as false-positive predictions of GIs if sequence composition bias (% GC content) were used as the only criterion to identify GIs. IslandViewer integrated the three most accurate GI prediction programs (Langille *et al.,* 2008; Waack *et al.,* 2006; Hsiao *et al.,* 2005), each using different approaches to predict GIs and, in effect, both HR-GIs were

detected by more than one tool. However, we looked for additional evidence that these gene-cassettes were not false positives, that is, that they were in a true genomic island. For this purpose, we built phylogenetic trees, with sequences from 20 Bacteroidetes genomes, of two genes found in HR1-GIs (*rpo*B and EF-Tu) and 16S rRNA. If these GIs were false positives, we would expect the phylogenies of rpoB and EF-Tu to match that of 16S rRNA. If these were true GIs subject to HR, however, we would expect somewhat different phylogenies for 16S rRNA on the one hand and *rpo*B and EF-Tu on the other (Fig. 3SM). Although the general topology among Flavobacteria, Sphingobacteria and Bacteroides branches was conserved with the three genes, several discrepancies could be detected between the 16S phylogeny on the one hand and those of the other two genes on the other (see black triangles and circles in Fig. 3SM). This is in accordance with the two functional genes following similar evolutionary trends and belonging to a GI.

A second line of evidence in support of the HR1 being true islands comes from a plasmid found in *Shewanella baltica* OS155 (pSbal03), containing the same gene-cassette of HR1-GI (except for one gene). The genes in this cassette were absent from the corresponding chromosome, further showing that HR1-GI is in fact laterally transferred (Fig. 6A). Interestingly, identical gene structure to this plasmid with a translocation of six ribosomal proteins was observed in 20 other *Shewanella* strains (Fig. 6A). Plasmid integration in the host chromosome by HR is a well known phenomenon in bacteria such *E.coli*, *Bacillus subtilis*, *Enterococcus faecalis* and others (Casey *et al*., 1991). Usually, the site of integration in the genome corresponds to the chromosome location of the fragment shared with the plasmid. In our case, the hypothetical insertion of the plasmid in the *Shewanella baltica* OS155 chromosome is located next to the chromosome EF-Tu gene shared by the plasmid and next to a large cluster of 15 ribosomal proteins and the *rpo*A (Fig. 6A). This cluster of ribosomal protein genes is considered to be a locally collinear block (LCB) meaning a contiguous segment of genes with low rearrangements (Dikow, 2011). The fact that most of the *Shewanella* strains harbor two copies of EF-Tu genes fits with the idea of one of them belongs to this or a similar *Shewanella* plasmid. We conducted phylogenetic reconstruction of EF-Tu genes in these 19 *Shewanella* strains and linked to the HR1-GI revealing certain anomalies in the tree topology (Fig. 4SM). For instance, both EF-Tu gene copies of *Shewanella* sp. MR-4 and MR-7, two isolates retrieved from different depths of the Black Sea (Venkateswaran *et al.,* 1999), clustered with each other instead of with the other EF-Tu gene copy of their genome (Fig. 6B). This finding suggests HR events and it is consistent with a recent study in which high level of HR has been discovered among co-occurring *Shewanella baltica* isolates (Caro-Quintero *et al.,* 2011). In summary, we are confident that this homologous recombination GIs are true genomic islands that in the case of S*hewanella* strains this HR1-GI was first transferred to

the genome by a plasmid and mobilized after by the site-specific recombinases within the GI. This finding is consistent with a recent study in which high level of HR has been discovered among co-occurring *Shewanella baltica* isolates (Caro-Quintero *et al*., 2011). It is known that recombination plays a cohesive role in bacteria within closely related lineages, because HR is rare between different taxa. (Caro-Quintero *et al.,* 2011, Fraser *et al.,* 2007, Fraser *et al,.*

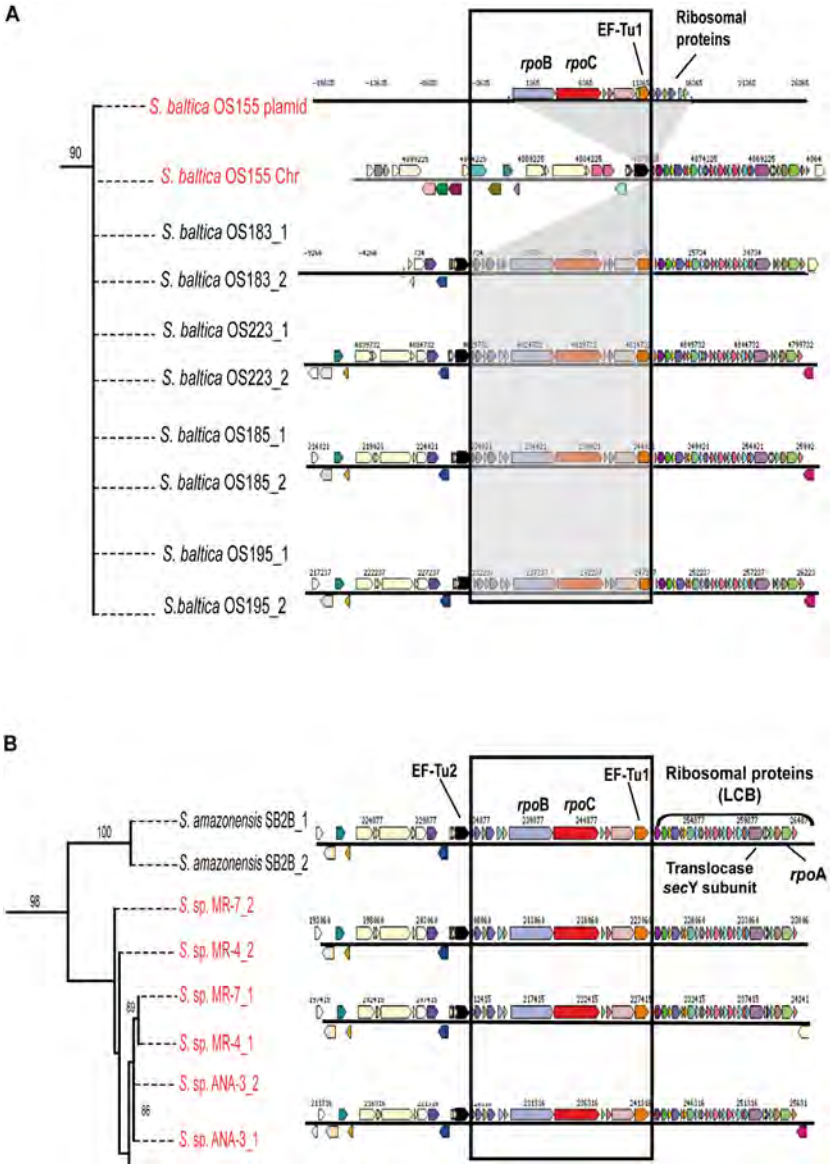


Figure 6. **Zoom of two sections of the phylogenic tree of the EF-Tu in *Shewanella* strains**. The phylogenetic tree show the two gene copies of the EF-Tu of 19 *Shewanella* strains linked to the hypothetical HR1-GI marked as a black box. **A)** The insertion location of the plasmid pSbal03 of *Shewanella baltica* OS155 in its chromosome (both labeled in red) is shown in grey and the most representative genes are indicated within HR1-GI. **B)** *Shewanella* strains labeled in red show discrepancies in the EF-Tu phylogeny when both EF-Tu genes were compared.

2009). However, HR and other mechanisms such genomic rearrangements have been also identified as a key role driving speciation in several aquatic bacterial populations. (Papke *et al.,* 2004, Zehr *et al*., 2007). We have observed identical HR-GIs not only within strains but also across genera in Bacteroidetes and this reinforces this latter possibility. In summary, we are confident that this homologous recombination GIs are true genomic islands that in the case of *Shewanella* strains this HR1-GI was first transferred to the genome by a plasmid and mobilized after by the site-specific recombinases within the GI.

Functional Annotation of the Prokaryotic Genomic Islands

One of the reported characteristic features of prokaryotic genomic islands is a higher concentration of genes coding for hypothetical proteins than in other genome regions (Hsiao *et al*., 2005). In effect, we found significantly higher percentage of hypothetical proteins within GIs than the average for the whole genome in 71% of the genomes (Fisher's test with the Bonferroni correction: $p < 0.05$) (Fig. 7). The 19 genomes with non-significant differences of HP within and outside the GIs were mostly marine Bacteroidetes or Cyanobacteria (Fig. 7). These two bacterial classes were the ones with largest % HP in their genomes (Fig. 7). We believe this is due to the lower number of well studied strains compared to Proteobacteria. Thus, the %HP is larger throughout the genome and therefore no significant differences were

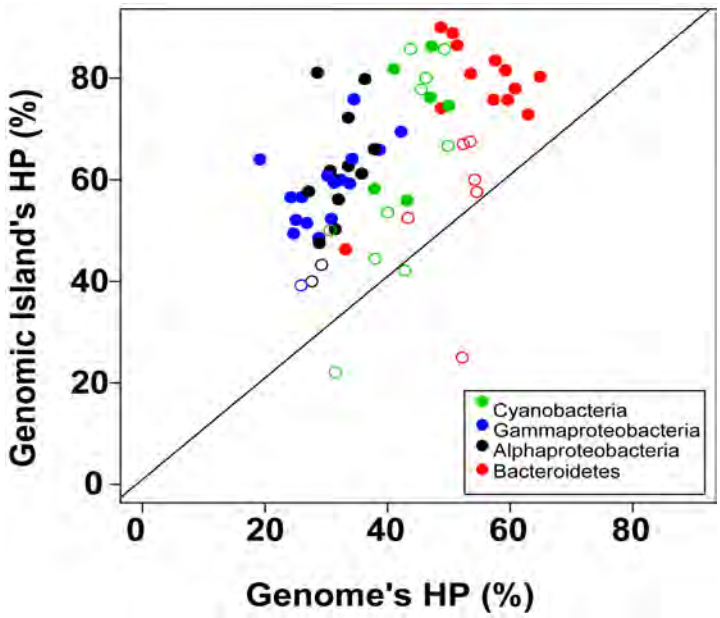


Figure 7. Percentage of hypothetical proteins (HP) within the GIs and on average for each bacterial genome. Each bacterial taxa is represented by different colors. Solid circles show statistically significant differences using the corrected P value (Bonferroni) < 0.05 after Fisher Exact Test, while empty circles did not show significant differences.

expected within and outside GIs. Around 55-60% of the genes in GIs coded for hypothetical proteins. In this respect, therefore, the GIs of marine bacteria are like those described before in pathogenic bacteria, where hypothetical proteins constituted 53% of the genes within GIs versus 28% in the rest of the genome (Hsiao *et al.,* 2005).

Next, we annotated the genes within GIs assigning them to functional categories with two approaches: Clusters of Orthologous Groups (COG) and GeneOntology (GO) with a total of 3725 and 3360 genes respectively to which a function could be assigned. Their distribution in the 22 COG categories appears in Figure 8. As expected category L (replication, recombination, and repair) contained over 20% of the total in agreement with the high proportion of transposases, integrases, and other recombinase related enzymes found in GIs. The next category was R, "general prediction only" with 11%. This basically includes proteins for which a more specific function could not be assigned and is, thus, not informative. Cell wall/membrane/envelope biogenesis (M) with 10% and the translation/ ribosomal structure and biogenesis (J) with 9% were especially well represented. Cell motility (N), defense mechanisms (E), and inorganic ion transport and metabolism (P) were also represented (3-4%).

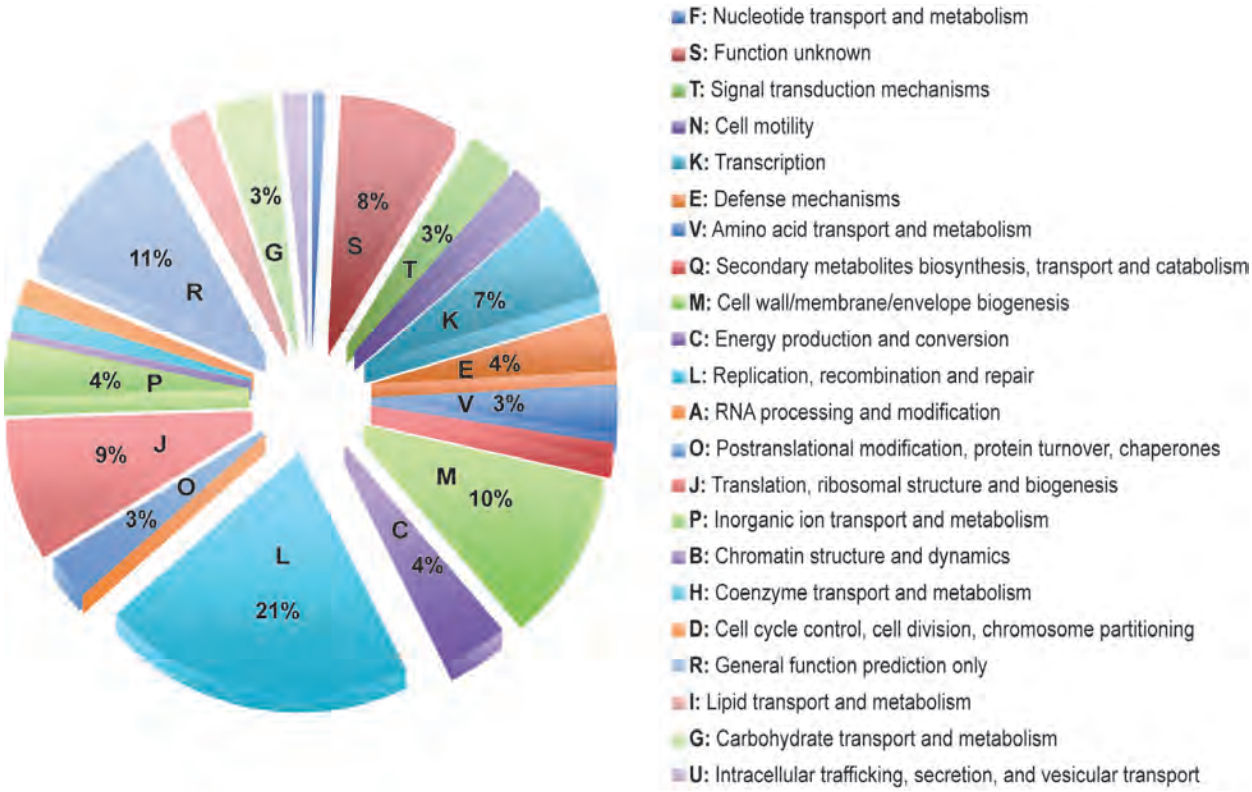


Figure 8. Distribution of annotated genes within the GIs according to their COG category. Percentage shown for those categories accounting for ≥3%.

In addition, we used Blast2GO to determine the functional annotation based on GO into the three major functional categories: Cellular Component (CC), Biological Process (BP) and Molecular Function (MF) (Fig. 9). CC category genes were very abundant especially those associated with the plasma membrane specifically (16%) or with membranes in general (36%). In the BP category, as expected, DNA integration (18%) and transposition DNA-mediated genes (14%) were very abundant. We found 13% of the genes were associated with translation. Other genes related to the two-component signal transduction system, and DNA repair and proteolysis related proteins with 3% each were frequent. In the MF category, the most abundant were the ATP binding (12%) and transposase activity (11%). Structural constituents of ribosomes (9%) including ribosomal proteins from small and large subunits (7 and 6%) were significant. Flagellin-based flagellum proteins (5%) were also frequent (Fig. 9). In summary, both functional annotation approaches revealed that a diverse range

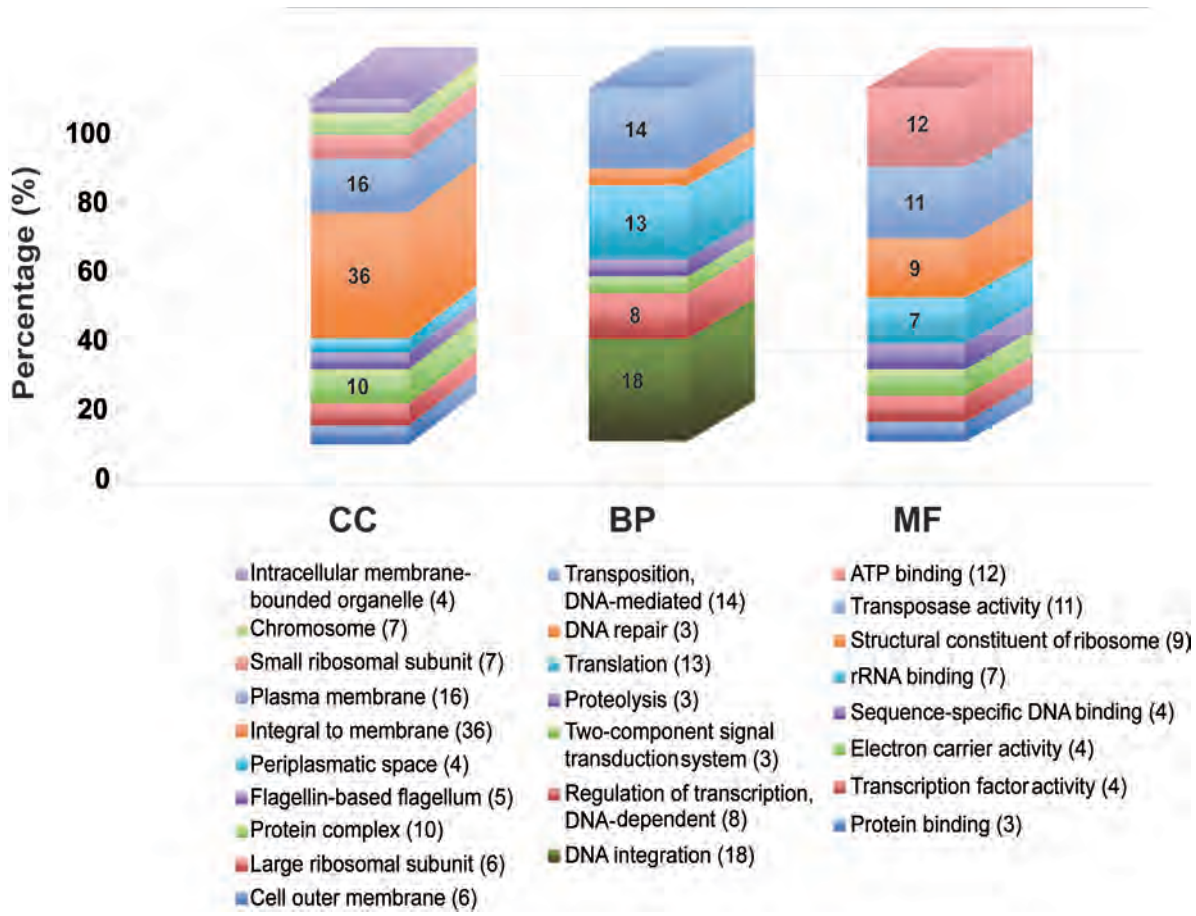


Figure 9. Distribution of annotated genes within the GIs according to gene ontology (GO) classification. Functional categories were split in three: Cellular Components (CC), Biological Processes (BP) Molecular Functions (MF). Numbers between parenthesis indicate the percentage of appearance (only shown if ≥3%).

of biologically relevant genes were present in the GIs. As expected, these included genes for mobility of DNA fragments but, interestingly, also genes associated with basic cellular mechanisms such as translation and regulation of transcription and transduction.

The possibility to move clusters of genes associated with transcription and their regulation (presence of β subunit of DNA-directed RNA polymerases, elongation factor Tu, sigma factors, transcription termination factors) and translation (within the mobility of ribosomal protein) between closely related genomes and/or different genera may be beneficial for bacterial fitness under changing environmental conditions when an increasing level of the transcription and synthesis of certain proteins may be needed. IS/Transposons transposases located nearby and within GIs may have a key role to activate transcription of those genes by either introducing complete or partial promoters located within the element itself, by disrupting another gene that may inhibit transcription (Syvanen 1984) or by inserting foreign genes into positions where they become regulated by endogenous promoters (Kasak *et al.,* 1993).

Differences in GI gene content among marine bacterial classes

To find out whether there were differences in the functional categories found within GIs of the four different bacterial classes, we compared the representation of GO terms in each main phylogenetic group with the remaining dataset by paired Fisher´s Exact Tests (Table 3SM). There were significant differences in all cases, indicating that each bacterial class had a different set of functions preferentially represented in their GIs. The GO terms that were significantly over-represented (p-value <0.01 after False Discovery Rate correction) in each of the four phylogenetic groups are listed in Table 2. Cyanobacteria were, by far, the taxon with highest and most diverse number of enriched GO terms (a total of 18). Most of these were related to photosynthesis, both to the antenna or photosystem proteins and to the electron transport system. These photosynthetic related genes (a total of 30 genes) were found within GIs in 50% of the Cyanobacteria genomes, suggesting that they are a rather common feature among Cyanobacteria GIs. Other GO terms enriched in cyanobacteria were linked to proteolysis/hydrolysis activity, glucose metabolism, histidine and cobalamin biosynthesis.

Alphaproteobacteria GIs were enriched (six GO terms) in genes related to transposases, DNA transposition activity, and locomotion. Surprisingly, RNA-directed DNA-polymerase activity (a signature of presence of prophages of retroviruses), and ribosome assemblage linked to a high number of ribosomal proteins, were also specifically enriched in marine

Alphaproteobacteria genomes. The two GO terms enriched in Gammaproteobacteria included genes associated to site-specific recombinases and ligase activity associated with DNA mobility and rearrangements.

Flavobacteria were specifically enriched in eight GO terms with genes associated to ATPase and GTPase activity and, interestingly, in processes related to DNA-directed RNA polymerases activity, transcription, and its regulation processes (Table 2). GIs of non-marine

Table 2. Gene Ontology (GO) terms enrichment analyses of GIs in 4 main phylogenetic groups.

Taxonomic Group	GO Term	Name	Ontology Category	FDR	FWER	Single test p-Value
Cyanobacteria	GO:0015979	Photosynthesis	BP	8.3E-7	4.1E-7	***
	GO:0030089	Phycobilisome	CC	3.0E-6	3.2E-6	***
	GO:0009898	Internal side of plasma membrane	CC	3.0E-6	3.7E-6	***
	GO:0018298	Protein-chromophore linkage	BP	7.0E-4	2.3E-3	***
	GO:0009521	Photosystem	CC	7.0E-4	2.3E-3	***
	GO:0046914	Transition metal ion binding	MF	1.7E-3	6.7E-3	***
	GO:0004175	Endopeptidase activity	MF	1.8E-3	8.5E-3	***
	GO:0006508	Proteolysis	BP	3.3E-3	3.0E-2	***
	GO:0016740	Transferase activity	MF	3.3E-3	3.0E-2	***
	GO:0009236	Cobalamin biosynthetic process	BP	3.4E-3	3.3E-2	***
	GO:0022900	Electron transport chain	BP	4.5E-3	5.0E-2	***
	GO:0006006	Glucose metabolic process	BP	5.0E-3	5.6E-2	***
	GO:0033178	Proton-transporting two-sector ATPase complex	CC	6.1E-3	9.3E-2	**
	GO:0042777	Plasma membrane ATP synthesis coupled proton transport	BP	6.1E-3	9.3E-2	**
	GO:0046933	Hydrogen ion transporting ATP synthase activity	MF	6.1E-3	9.3E-2	**
	GO:0000105	Histidine biosynthetic process	BP	6.1E-3	9.3E-2	**
	GO:0043231	Intracellular membrane-bounded organelle	MF	6.1E-3	9.8E-2	**
	GO:0016820	Hydrolase activity, acting on acid anhydrides	MF	7.2E-3	1.2E-1	**
Gammaprotebacteria	GO:0009009	Site-specific recombinase activity	MF	7.3E-4	1.8E-4	***
	GO:0016874	Ligase activity	MF	7.4E-4	3.7E-4	***
Alphaproteobacteria	GO:0042255	Ribosome assembly	BP	3.3E-9	5.0E-9	***
	GO:0003964	RNA-directed DNA polymerase activity	MF	3.6E-4	9.1E-4	***
	GO:0003995	Acyl-CoA dehydrogenase activity	MF	4.1E-4	1.1E-3	***
	GO:0040011	Locomotion	BP	7.6E-4	2.3E-3	***
	GO:0004803	Transposase activity	MF	1.2E-3	3.8E-3	***
	GO:0006313	Transposition, DNA-mediated	BP	5.5E-3	2.0E-2	***
Flavobacteria	GO:0015662	ATPase activity, coupled to transmembrane movement of ions	MF	1.8E-3	1.4E-3	***
	GO:0003899	DNA-directed RNA polymerase activity	MF	2.2E-3	2.9E-3	***
	GO:0003711	Transcription elongation regulator activity	MF	2.2E-3	1.0E-2	***
	GO:0032968	Positive regulation of RNA elongation from RNA pol. II promoter	BP	2.2E-3	1.0E-2	***
	GO:0003924	GTPase activity	MF	3.5E-3	1.8E-2	***
	GO:0032549	Ribonucleoside binding	MF	3.7E-3	2.4E-2	***
	GO:0031564	Transcription antitermination	BP	3.7E-3	2.4E-2	***
	GO:0008135	Translation factor activity, nucleic acid binding	MF	5.0E-3	3.6E-2	***
Non-	GO:0003735	Structural constituent of ribosome	MF	1.3E-9	5.3E-9	***
Marine Bacteroidetes	GO:0006412	Translation	BP	1.3E-9	5.3E-9	***
	GO:0019843	rRNA binding	MF	1.3E-9	5.3E-9	***
	GO:0015935	Small ribosomal subunit	CC	3.1E-7	2.1E-6	***
	GO:0000049	tRNA binding	MF	1.5E-6	1.1E-5	***
	GO:0015934	Large ribosomal subunit	CC	1.8E-3	1.4E-2	***
	GO:0003917	DNA topoisomerase type I activity	MF	5.3E-3	4.3E-2	***

Fisher Exact Test was used with distinct statistically methods: False Discovery Rate control (FDR), Family Wise Error Rate (FWER) and the p-Value without multiple testing corrections (single Test p-Value). Only the most specific GO terms over-represented in each bacterial group using FDR with statistical significance (**p-value <0.01; ***p-value <0.001) are shown. Five bacterial groups were analyzed: Cyanobacteria, Gammaproteobacteria, Alphaproteobacteria; Flavobacteria; *and* non-marine Bacteroidetes.

Bacteroidetes were enriched in seven GO terms, entirely linked to translation processes and involved in the structure of the ribosome (many ribosomal proteins associated with the large and small ribosomal subunits) and rRNA and tRNA binding.

Therefore, each bacterial class has a different set of genes in their GIs, indicating a different strategy in the role played by GIs in their ecology. This will be analyzed in the next section.

Biologically relevant genes within Marine Bacterial GIs

To analyze the ecological relevance of the genes found within GIs, we assigned them to 16 biological categories with the potential to increase bacterial fitness (see the complete list in Material and Methods). In Figure 10 we show a summary of the numbers of genes associated to each category in each genome analyzed (histograms in Figure 10). We found that many of the biological categories were widely distributed among the bacterial taxa, such as energy metabolism (number 2 in the histogram legend of Fig. 10), ribosomal proteins (3), hydrolysis activity (4), DNA restriction modification systems (6), β-subunit of DNA-directed RNA polymerase (7), transporters (8), two component systems (9), stress response proteins (11) or MGE (12). Such categories were well distributed among the genomes but exhibiting differences in their abundance within the GIs.

Transposases and integrases (category 12 in Fig. 10) can modify the structure of the genome through the transfer of DNA sequences to new locations within or between genomes (Rice and Baker 2001). Recently, it has been reported that transposases are the most abundant and ubiquitous genes in nature (Aziz *et al.*, 2010). We found a total of 675 genes related to transposases such as IS elements or transposons (Tn7/Tn27) representing about 8.2% of the total database. These genes were over-represented in Alphaproteobacteria within genomic islands (see significance test in Table 2) with almost a 30% of total transposases (194 genes) within this group (data not shown). Also, ribosomal proteins (category 3 in Fig. 10) accounted for 3% of all genes in our dataset (253 out of 8152 genes) and were very abundant in Alphaproteobacteria and Gammaproteobacteria genomes (also in non-marine Bacteroidetes) with almost a 25% and 22% respectively (data not shown). Specifically, marine genomes with more than 10 gene copies were *Sulfitobacter* sp. EE36 (32 copies) and *Silicibacter pomeroyi* DSS-3 (18 copies) in the Alphaproteobacteria, and S*hewanella baltica* OS155 (26 copies) and *Vibrio cholera*e O395 (29 copies) in the Gammaproteobacteria. Similarly, virulence gene clusters (category 14 in Figure 10) were found in all main taxa except for Alphaproteobacteria with the highest number for *Vibrio cholerae* O395 with 19

copies corresponding to the well known TCP (Toxin-Coregulated Pilus) located in one of the pathogenic island organized as a prophage (Karaolis *et al*., 1999). Finally, the presence of secretion system proteins (category 16 in Fig. 10) was found in all taxa but for marine Flavobacteria. Specifically, *Thalassiobium* sp. R2A62 exhibited seven copies of Type I secretion system proteins and at least 3 copies of Type III secretion system were found in *Shewanella baltica* OS155. These virulence associated secretion system proteins have been described in other marine genomes. Particularly, Type IV and Type VI secretions system genes were recurrent for Alpha- and Gammaproteobacteria genomes respectively (Persson *et al*., 2009).

Our GI dataset also included genes involved in protection from bacteriaophages such as DNA modification restriction systems (Type I, II and III, category 6 in Fig. 10). These systems are sequence-specific restriction enzymes, also called restriction endonucleases that have been known for a long time to act as protection from foreign DNA, such as bacteriophages (Werner Arber and Linn 1969). We detected 99 genes linked to restriction modification systems spread along all taxa but specially represented in: *Anabaena variabilis* ATCC29413 with five copies, *Roseobacter denitrifican*s Och 114 (10 copies), *Silicibacter pomeroy*i DSS-3 (eight copies) and *Psychrobacter cryohalolentis* K5 with 11 of them. Finally three marine Bacteroidetes had five copies each (*Cytophaga hutchinsonii* ATCC 33406, *Kordia algicida* OT-1 and *Robiginitalea biformata* HTCC2501) (Fig. 10). Interestingly, most of the restriction-modification system genes in our GI dataset were Type I (the most complex) with at least 36 genes, but we also found some representatives of type II and III (data not shown).

Another mechanism proposed to confer resistance from phage and possibly from other mobile elements is the presence of clustered regularly interspaced short palindromic repeats (CRISPRs) with their associated cas genes (Mojica *et al*., 2000; Barrangou *et al*., 2007; Sorek *et al*., 2008). These genetics elements have been identified in approximately 40% and 90% of Bacteria and Archaea genomes, respectively (Grissa *et al*., 2007). Recent research has shown that CRISPR systems could be primarily transferred by horizontal gene transfer and can be found over-represented within GIs (Ho Sui *et al*., 2009). Although we did not find many of them (category 13 in Figure 10) associated to our marine prokaryotic GI dataset, we found cas genes in Anabaena variabilis (four genes), in *Rhodobacter sphaeroides* KD131 (seven genes) and four genes in *Leeuwenhoekiella blandensis* MED217. Some of these CRISPRs systems were already in the CRISPRdb (http://crispr.u-psud.fr/crispr/) in the chromosome of these genomes although they were not associated with GIs and no CRISPRs had been previously identified in *Leeuwenhoekiella blandensis* MED217.

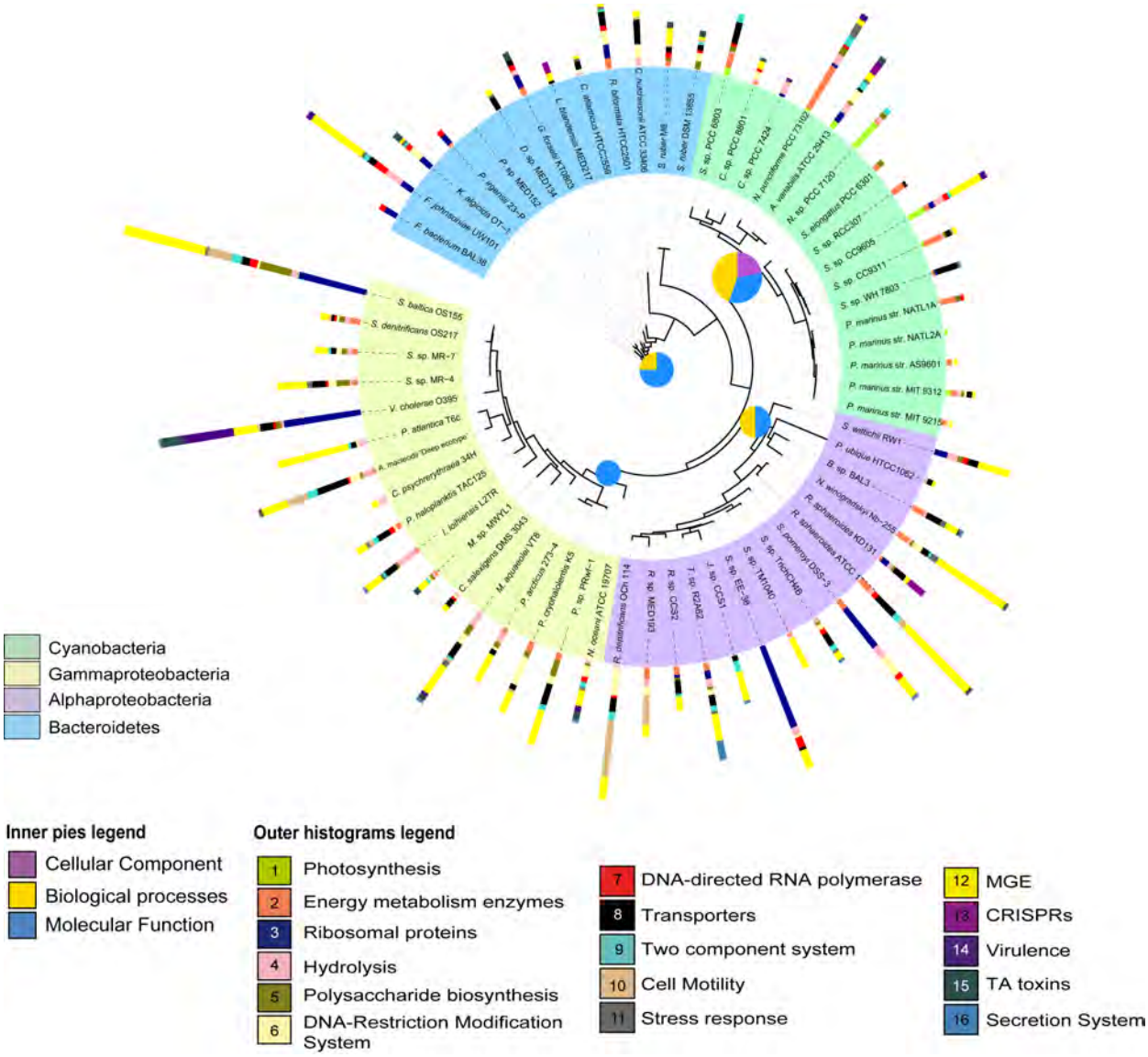


Figure 10. **Phylogenetic representation of the 66 bacterial genomes with genomic islands**. The colored ring shows the four main phylogenetic groups analyzed; the histograms indicate the number of genes (in absolute number) within genomic islands associated with any of the 16 biological categories described. The inner pies shown the functional category specifically over-represented for Cyanobacteria, Gammaproteobacteria, Alphaproteobacteria and Flavobacteria wherein pie size is proportional to the number of GO functional categories enriched.

Other categories were restricted to a few genomes and/or phylogenetic groups. This was the case for photosynthetic genes (category 1 in Figure 10) within cyanobacterial taxa. Photosystem I subunits and photosynthesis antenna proteins as well as electron transporter systems (ATP synthases) or ferredoxins were detected in Cyanobacteria genomic islands. Photosynthesis genes within genomic islands have been well described for *Procholorococcus*

and *Synechococcus* strains (Coleman *et al.*, 2006; Kettler *et al.*, 2007; Dufresne *et al.*, 2008) but we also found them in *Synechocystis* sp. PCC 6803 with three phycobilisome linker proteins in the 14.6 kb GI, as well as in *Anabaena variabilis* ATCC29413 with two photosystem I subunit proteins and one ferredoxin, and in *Nostoc* sp. PCC7120 with eight ATP synthase subunits, two phycobilisome linker proteins and another two allophycocyanin alpha/beta subunits.

These photosynthetic related genes are linked to relevant physiological characteristics of these photoautotrophic bacteria. The acquisition of these genes by GIs may provide specific light niche adaptations to specific strains (Dufresne *et al.*, 2008), underlining the need for analyzing their genomic islands to fully understand the ecology of Cyanobacteria.

In addition, we detected cell motility (flagella) genes (category 10 in Fig. 10) constrained basically to *Rosebacter denitrificans* OCh 114 and *Roseobacter* sp. MED193 with 20 and 11 copies respectively, and *Shewanella baltica* OS155 with eight. Cell motility by flagella is a very important physiological characteristic to allow bacteria to move towards favorable environmental conditions, form biofilms and/or acquire nutrients. Genes to reconstruct the flagellum structure can include more than 50 but only 24 are considered as the core set and are present in most flagellated bacterial taxa (Liu and Ochman 2007). Some of theses genes can be acquired through HGT events as described in *Photobacterium profundum* SS9 (Campanaro *et al.*, 2005) wherein a cluster of genes involved in the lateral flagellar synthesis was present in the GI and absent in a closely related strain and thus has been horizontally transferred. Our dataset agrees with this view with a set of 43 genes within two *Roseobacter* genomes and two Gammaproteobacteria. Genes related to the flagellar basal rod, body, ring, and hook or flagellin proteins were repeatedly found in these GIs. In R*oseobacter denitrificans* OCh 114, 14 of these flagellar genes were concentrated in two GIs of 14.1 kb and 16 kb with 12 and 8 genes respectively. These GIs were flanked by transposases, integrases or phage-related integrases while in *Roseobacter* sp. MED193 these genes were in a single GI of 12 kb, also flanked by phage-integrase genes. In addition, *Shewanella baltica* OS155 had eight flagellar related genes in the GI of 11.5 kb and interestingly, a chemotaxis protein was located within the same GI. Moreover, one of the GIs (18.7 kb) of *Alteromonas macleodii* "Deep ecotype" displayed eight flagellar proteins. This GIs was the GI8 (20 kb) previously reported by Ivars-Martínez *et al.* (2008). We investigated whether these flagellar proteins present in our GIs dataset were also present in the genome but we found different genes for the flagellum structure in the chromosome (data not shown). Finally, conjugative transposons (category 12 in Fig. 10) were found only in a few taxa. Two Alphaproteobacteria, *Jannaschia* sp. CCS1 and *Sphingomonas wittichii* RW1 displayed six and seven genes related to these conjugative transposons, and the Bacteroidetes *Flavobacterium johnsoniae* UW101

had four copies.

It is highly probable that many genes within GIs are positively selected due to the potential benefits of these genes for the bacterial life-style. It has been shown that GIs of *Prochlorococcus* strains displayed differential expression under light and nutrient stress conditions (Coleman *et al.,* 2006). Also, *Prochlorococcus* GIs contain genes related to the viruses attachment to the host cell surface and those GIs have an important role to the viruses-host coexistence (Avrani *et al.*, 2011). Moreover, one of the GIs with heavy metal resistance genes detected in *Alteromonas macleodii* "Deep ecotype" conferred to this strain higher resistance to mercury and zinc concentrations (Ivars-Martínez *et al.*, 2008). In a different context, in the Actinobacteria *Salinispora arenicola*, orthologs within GIs showed evidence of positive selection compared with the non-island genes (7.6% vs. 1.6%) (Penn *et al.*, 2009). Also, the hyperhalophilic bacterium *Salinibacter ruber* strains M8 and M31 displayed 40 strains-specific genes present in their GIs with a ratio of substitution rates at non-synonymous and synonymous sites >1 (dN/dS >1) in which 25 of them were hypothetical proteins (Peña *et al.,* 2009). Consequently, there is some evidence for the adaptive significance of genomic island genes among environmental bacteria, but the extent and effect on diversification mechanisms in marine bacterial taxa is still unknown.

CONCLUSIONS

Genomic islands were present in most of the marine bacterial genomes analyzed. Our results indicated that both horizontal gene transfer (HGT) by phages, plasmids and MGE and homologous recombination (HR) play an important role for the mobility of clusters of genes between taxa and within closely related genomes, thus modulating the flexible pool of the genome. Our findings provide insights on the fact that GIs may increase bacterial fitness under changing environmental conditions by providing, not only novel foreign genes, but also modulating their transcription, regulation, and/or transduction. The potential role that genomic islands have in rearranging the structure, and increasing the diversity, of marine bacterial genomes is emphasized by the results presented here. We observed that some genomic islands were intimately associated with the physiology and ecology of the microorganisms but we also found some relevant housekeeping genes in theory linked to core genome. These results would reinforce the need to establish a pangenome concept for marine bacterial species wherein genomic islands would be crucial to fully understand the ecology and evolution of marine bacterial in the ocean. Exploring the mechanisms maintaining and selecting GIs is the next logical step to gain insights into the evolutionary processes shaping marine bacterial genomes.

ACKNOWLEDGEMENTS

We thank Guillem Salazar for his help on statistical analyses of the data and Ramiro Logares for critical reading.
BF-G was a recipient of a I3P grant from CSIC and AF-G was supported by grant CONSOLIDER-INGENIO2010 GRACCIE CSD2007-00067 from the Spanish Ministry of Science and Innovation (MICINN) to EOC. SGA was supported by a RyC contract from Spanish Ministry of Science and Innovation. This research was supported mostly by grant MICRODIVERSITY (CGL2008-00762/BOS) to SGA but also CGL2009-13318-CO2-01/BOS to EOC from MICINN.

REFERENCES

Alonso C, Warnecke F, Amann R, Pernthaler J. (2007). High local and global diversity of Flavobacteria in marine plankton. *Environ Microbiol* **9:** 1253-1266.

Alonso-Sáez L, Gasol J. (2007). Seasonal variations in the contributions of different bacterial groups to the uptake of low-molecular-weight compounds in Northwestern Mediterranean Coastal waters. *Appl Environ Microbiol* **73:** 3528-3535.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* **215:** 403-410.

Amaral-Zettler L, Artigas LF, Baross J, Loka Barathi PA, Boetius A, Chandramohan D *et al.,* (2010). A Global Census of Marine Life. In: Alasdair D. McIntyre, Ed. Life in the World's Oceans. Blackwell Publishing Ltd., pp 223-245.

Bauer M, Kube M, Teeling H, Richter M, Lombardot T, Allers E *et al.,* (2006). Whole genome analysis of the marine Bacteroidetes *Gramella forsetii* reveals adaptations to degradation of polymeric organic matter. *Environ Microbiol* **8:** 2201-2213.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340: 783-795.

Castresana J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17:** 540-552.

Cottrell MT, Kirchman DL. (2000). Natural assemblages of marine proteobacteria and members of the Cytophaga-Flavobacteria Cluster consuming low- and high-molecular-weight dissolved organic matter. *Appl Environ Microbiol* **66:** 1692-1697.

DeLong EF, Franks DG, Alldredge AL. (1993). Phylogenetic diversity of aggregate-attached vs. free-fiving marine macterial assemblages. *Limnol Oceanogr* **38:** 924-934.

DeLong EF, Béjà O. (2010). The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times. *PLoS Biol* **8:** e1000359.

Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz H-R *et al.,* (2008). The Pfam protein families database. *Nucleic Acids Res* **36:** D281-D288.

Fuhrman JA, Schwalbach MS, Stingl. (2008). Proteorhodopsins: an array of physiological roles? *Nat Rev Micro* **6:** 488-494.

Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D *et al.,* (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309:** 1242-1245.

Glöckner FO, Fuchs BM, Amann R. (1999). Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence in situ hybridization. *Appl Environ Microbiol* **65:** 3721-3726.

Goldberg SMD, Johnson J, Busam D, Feldblyum T, Ferriera S, Friedman R *et al.,* (2006). A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. *Proc Natl Acad Sci USA* **103**:11240-11245.

Gómez-Consarnau L, González JM, Coll-Lladó M, Gourdon P, Pascher T, Neutze R *et al.,* (2007). Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. *Nature* **445:** 210-213.

Gómez-Consarnau L, Akram N, Lindell K, Pedersen A, Neutze R, Milton DL *et al.,* (2010). Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. *PLoS Biol* **8:** e1000358.

Gónzalez JM, Fernández-Gómez B, Fernàndez-Guerra A, Gómez-Consarnau L, Sánchez O, Coll-Lladó M *et al.,* (2008). Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (Flavobacteria). *Proc Natl Acad Sci USA* **105**:8724-8729.

González JM, Pinhassi J, Fernández-Gómez B, Coll-Llado M, González-Velázquez M, Puigbó P *et al.,* (2011). Genome analysis of *Dokdoni*a sp. MED134 highlights the convergent evolution of proteorhodopsin-containing marine Bacteroidetes. *Environ Microbiol* **77:**8676-8687.

Haft DH, Selengut J, Mongodin EF, Nelson KE. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* **1:** e60.

Hou S, Saw JH, Lee KS, Freitas TA, Belisle C, Kawarabayasi Y *et al.,* (2004). Genome sequence of the deep-sea gamma-proteobacterium *Idiomarina loihiensis* reveals amino acid fermentation as a source of carbon and energy. *Proc Natl Acad Sci USA* **101:** 18036-18041.

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D *et al.,* (2009). InterPro: the integrative protein signature database. *Nucleic Acids Res* **37:** D211-D215.

Jarrell KF, McBride MJ. (2008). The surprisingly diverse ways that prokaryotes move. *Nat Rev Micro* **6:** 466-476.

Kimura H, Young CR, Martínez A, DeLong EF. (2011). Light-induced transcriptional responses associatd with proteorhodopsin-enhanced growth i marine flavobacterium. *ISMEJ* **5:** 1641–1651

Kirchman D. (2002). The ecology of Cytophaga-Flavobacteria in aquatic environments. *FEMS Microbiol Ecol* **39:** 91-100.

Krogh A, Larsson Br, von Heijne G, Sonnhammer EL. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol* **305:** 567-580.

Langille MGI, Brinkman FSL. (2009). IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* **25:** 664-665.

Langille MGI, Hsiao WWL, Brinkman FSL. (2010). Detecting genomic islands using bioinformatics approaches. *Nat Rev Micro* **8:** 373-382.

Letunic I, Bork P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23:** 127-128.

Letunic I, Bork P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* **39**: W475-W478.

Martens EC, Koropatkin NM, Smith TJ, Gordon JI. (2009). Complex Glycan Catabolism by the Human Gut Microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem* **284:** 24673-24677.

McBride MJ, Xie G, Martens EC, Lapidus A, Henrissat B, Rhodes RG *et al.,* (2009). Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. *Appl Environ Microbiol* **75:** 6864-6875.

Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J *et al.,* (2003). GenDB--an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res* **31:** 2187-2195.

Pinhassi J, Sala MM, Havskum H, Peters F, Guadayol O, Malits A *et al.,* (2004). Changes in bacterioplankton composition under different phytoplankton regimens. *Appl Environ Microbiol* **70:** 6753-6766.

Pommier T, Canback B, Riemann L, Bostrom KH, Simu K, Lundberg P *et al.,* (2007). Global patterns of diversity and community structure in marine bacterioplankton. *Mol Ecol* **16:** 867-880.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies Jr *et al.,* (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35:** 7188-7196.

Pushker R, Mira A, Rodriguez-Valera F (2004). Comparative genomics of gene-family size in closely related bacteria. *Genome Biol* **5:** R27.

Qin Q-L, Zhang X-Y, Wang X-M, Liu G-M, Chen X-L, Xie B-B *et al.,* (2010). The complete genome of *Zunongwangia profunda* SM-A87 reveals its adaptation to the deep-sea environment and ecological role in sedimentary organic nitrogen degradation. *BMC Genomics* **11:** 247.

Richter M, Lombardot T, Kostadinov I, Kottmann R, Duhaime M, Peplies J *et al.,* (2008). JCoast - A biologist-centric software tool for data mining and comparison of prokaryotic (meta) genomes. *BMC Bioinformatics* **9:** 177.

Schut F, de Vries EJ, Gottschal JC, Robertson BR, Harder W, Prins RA *et al.,* (1993). Isolation of typical marine bacteria by dilution culture: growth, maintenance, and characteristics of isolates under laboratory conditions. *Appl Environ Microbiol* **59:** 2150-2160.

Stamatakis A, Hoover P, Rougemont J. (2008). A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol* **57:** 758-771.

Steindler L, Schwalbach MS, Smith DP, Chan F, Giovannoni SJ. (2011). Energy starved Candidatus *Pelagibacter ubique* substitutes light-mediated ATP production for endogenous carbon respiration. *PLoS ONE* **6:** e19725.

Stingl U, Desiderio RA, Cho J-C, Vergin KL, Giovannoni SJ. (2007). The SAR92 Clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. *Appl Environ Microbiol* **73:** 2290-2296.

Tatusov RL, Koonin EV, Lipman DJ. (1997). A Genomic perspective on protein Ffamilies. *Science* **278:** 631-637.

Woyke T, Xie G, Copeland A, González JM, Han C, Kiss H *et al.,* (2009). Assembling the marine metagenome, one cell at a time. *PLoS ONE* **4:** e5299.

Yooseph S, Nealson KH, Rusch DB, McCrow JP, Dupont CL, Kim M *et al.,* (2010). Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* **468:** 60-66.

SUPPORTING INFORMATION

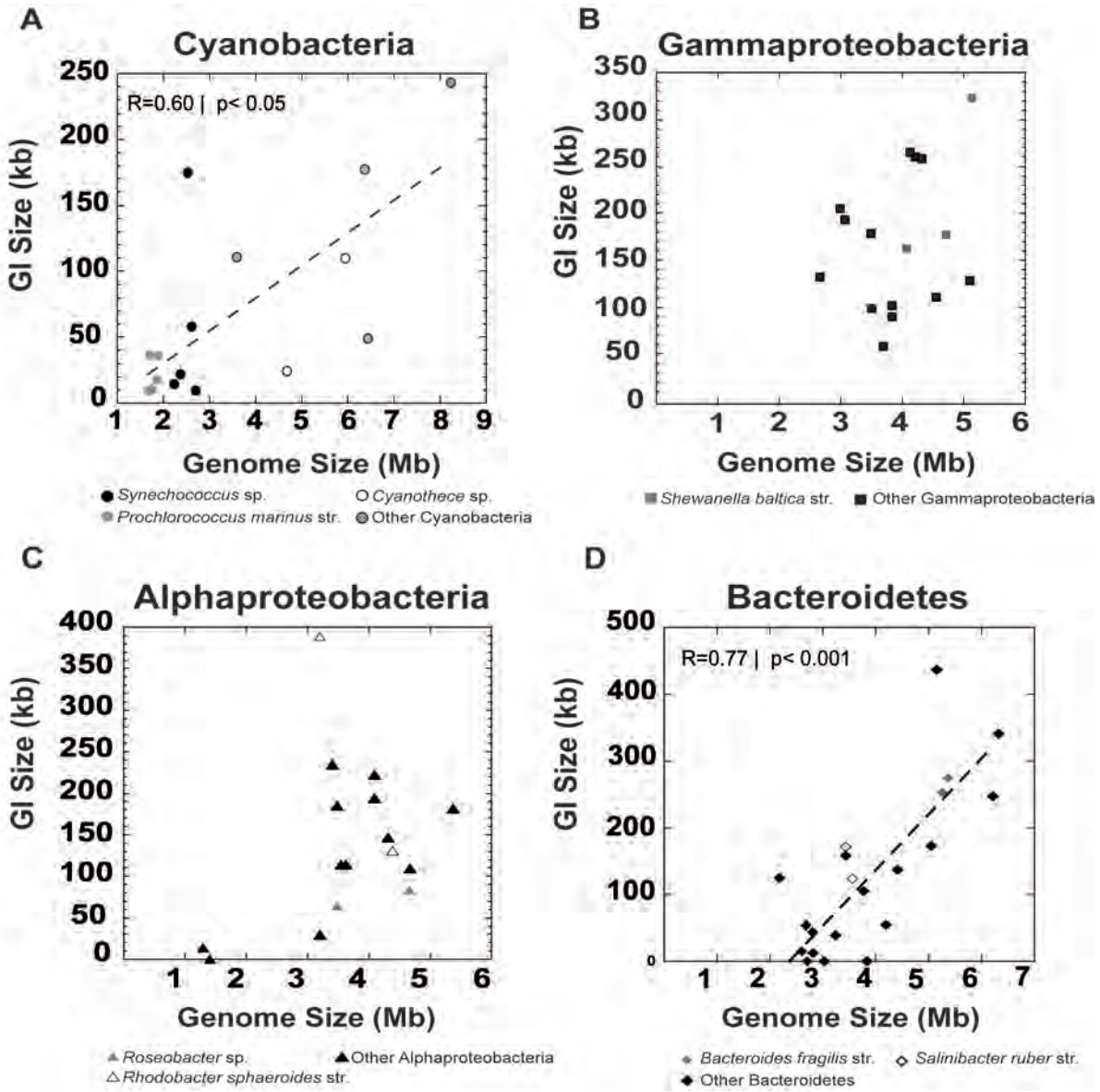


Figure 1SM. **Patterns of GIs in four main phylogenetic groups.** Relationship between total GI size (kb) and genome size (Mb) in **A**, Cyanobacteria; **B**, Gammaproteobacteria; **C**, Alphaproteobacteria; and **D**, Bacteroidetes. Discontinuous lines in A and D indicate regressions that were marginally significant with R values of 0.60 and 0.77 respectively.

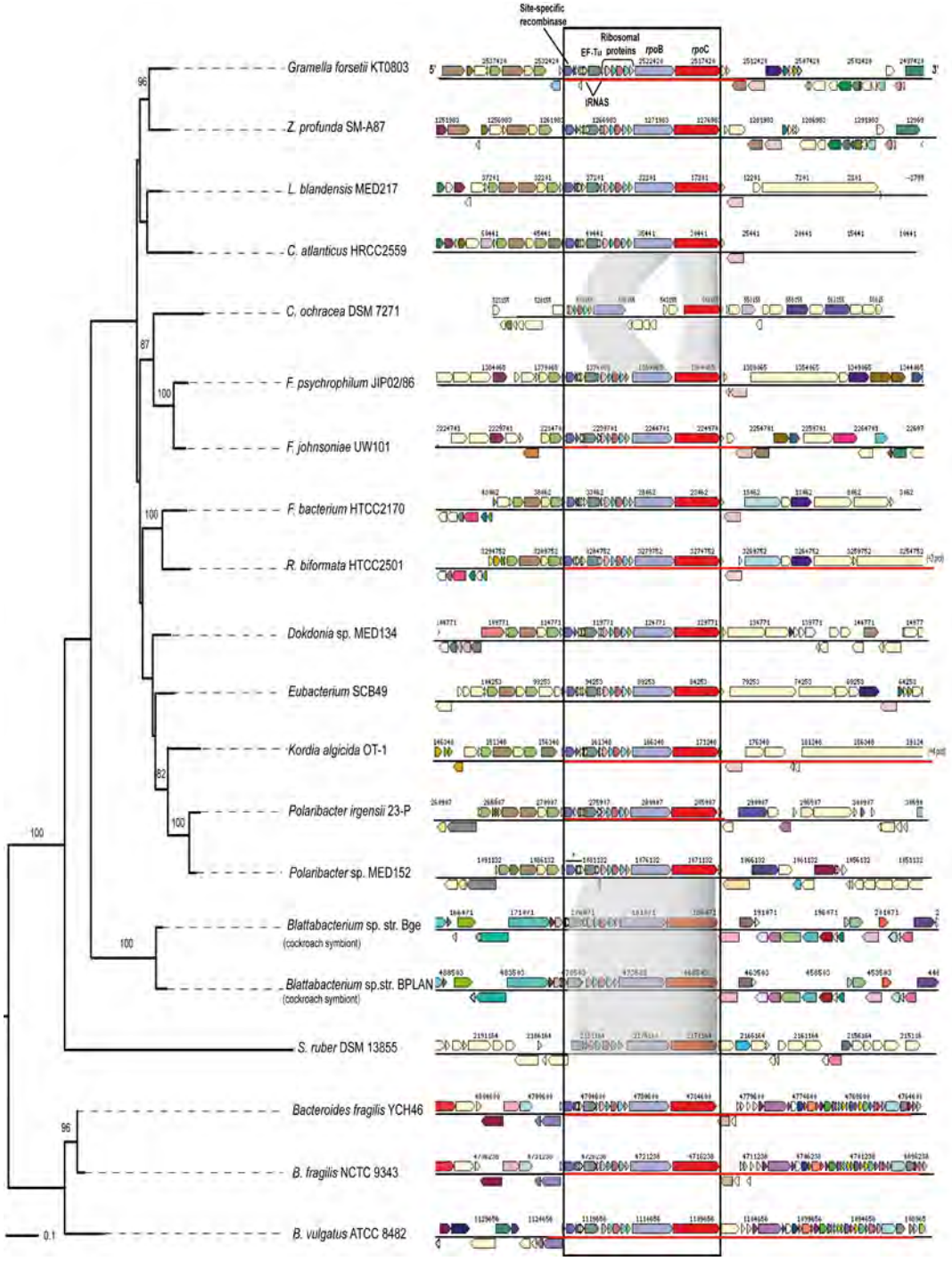


Figure 2SM. Phylogeny of the 16S rRNA gene of 20 Bacteriodetes genomes linked to the HR1-GI. Conserved gene cassette within HR1-GI is represented as a black rectangle. Red underline indicates the length of the HR1-GIs detected in genomes from our dataset. Genomes shaded indicated either insertion or deletion in the corresponding genomes. The genes that were absent in Sphingobacteria genomes are indicated with a black line and an asterisk in the *Polaribacter* sp. MED152 as a genome reference with presence of these genes.

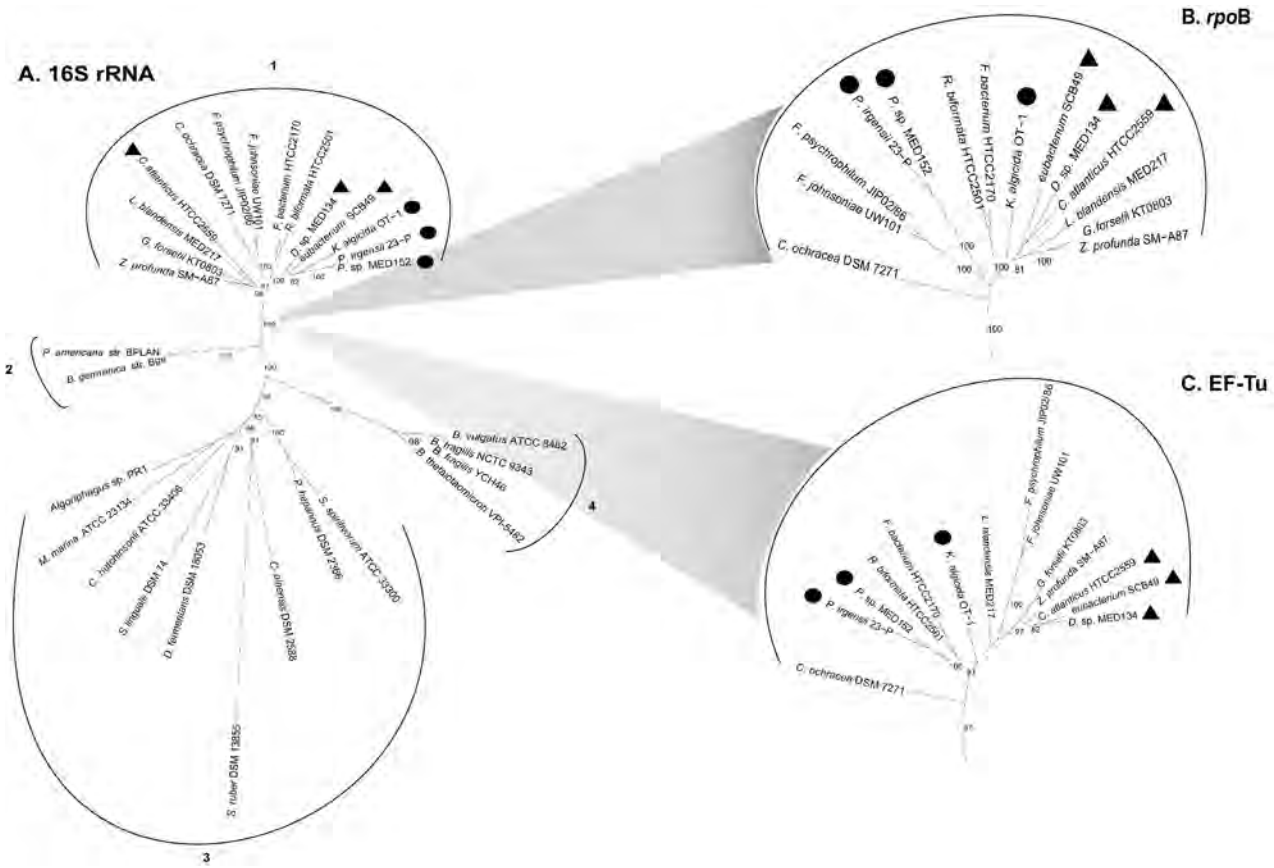


Figure 3SM. Phylogeny comparison of the 16S rRNA, EF-Tu and *rpo*B genes in 20 Bacteroidetes genomes. **(A)** 16S rRNA phylogenetic tree, numbers represent the main clusters that are well maintained along the three phylogenies. The two zoom sections display the *rpo*B **(B)** and EF-Tu **(C)** phylogenies wherein different symbols (black circles and triangles) represent specific taxa within marine flavobacteria with discrepancies between 16S rRNA and *rpo*B/ EF-Tu phylogenies.

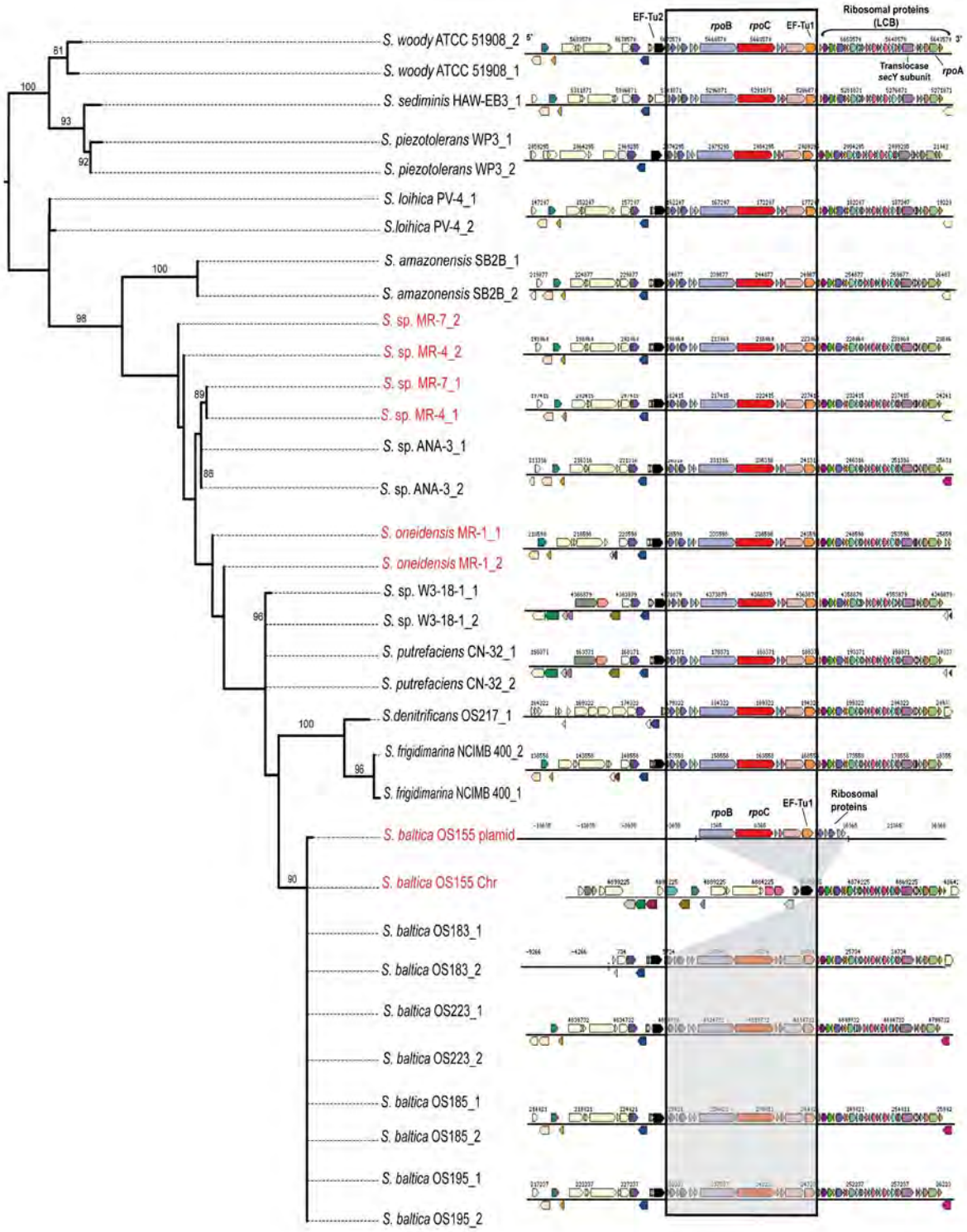


Figure 4SM. Phylogeny EF-Tu linked to their HR-GI within 19 *Shewanella* strains. Conserved gene cassette within HR1-GI is represented as a black rectangle. Two zooms of the phylogenetic sections are represented in Figure 6.

Table 1SM. Total number of genes and genes where function could be assigned for the manually and automatically annotated genomes.

	Manually annotated GIs		Automatically annotated GIs		Fisher´s Test
Control genome	**Total genes**	**Genes with known function[a]**	**Total genes**	**Genes with known function[a]**	**Significant difference**
Prochlorococcus marinus str. MIT9312	239	82	44	16	NO
Synechococcus sp. WH7803	423	154	75	26	NO
Synechococcus sp. CC9311	705	225	128	51	NO
Synechococcus sp. CC9605	565	159	323	79	NO
Synechococcus sp. RCC307	448	138	56	28	NO
Alteromonas macleodii "Deep ecotype"	467	190	276	116	NO
Salinibacter ruber M8	282	51	171	33	NO
Salinibacter ruber DSM13855	158	66	101	48	NO

Significance (p-value <0.01) of Fisher´s exact test was used to detect significant differences between them.

[a] The difference with total genes is accounted for by hypothetical proteins.

Table 2SM. List of the 70 bacterial genomes used for our analyses with the number of genomic islands detected, total GI size (in kb) and the percentage of the bacterial genome represented by GIs.

Bacterial genomes	Genome Size (Mb)	# of GIs (≥9.5 kb)[a]	GIs Size (kb)	Ratio* (%)
Cyanobacteria (16)				
Anabaena variabilis ATCC 29413	6.36	9	177.61	2.79
Cyanothece sp. PCC 7424	5.94	4	110.28	1.86
Cyanothece sp. PCC8801	4.67	2	24.65	0.53
Nostoc punctiforme PCC 73102	8.23	12	243.35	2.96
Nostoc sp. PCC 7120	6.43	3	49.01	0.76
Prochlorococcus marinus str. AS9601	1.67	1	9.99	0.6
Prochlorococcus marinus str. MIT 9215	1.73	1	10.47	0.61
Prochlorococcus marinus str. MIT 9312	1.7	3	36.93	2.17
Prochlorococcus marinus str. NATL2A	1.84	1	18.18	0.99
Prochlorococcus marinus str. NATL1A	1.86	3	36.63	1.97
Synechococcus elongatus PCC 6301	2.69	1	10.14	0.38
Synechococcus sp. CC9311	2.6	2	58.37	2.24
Synechococcus sp. CC9605	2.51	11	175.45	6.99
Synechococcus sp. RCC307	2.22	1	14.87	0.67
Synechococcus sp. WH 7803	2.36	2	21.87	0.93
Synechocystis sp. PCC 6803	3.57	5	110.91	3.11
Gammaproteobacteria (17)				
Alteronomas macleodii 'Deep ecotype'	4.2	10	260.47	6.20
Chromohalobacter salexigens DMS 3043	3.69	4	57.97	1.57
Colwellia pshychrerythraea 34H	5.37	3	72.81	1.36
Idiomarina loihiensis L2TR	3.83	5	90.13	2.35
Marinobacter aquaeoloi VT8	4.32	17	258.13	5.98
Marinomonas sp. MWYL1	5.1	5	127.88	2.51
Nitrosococcus oceani ATCC 19707	3.48	7	178.25	5.12
Pseudoalteronomas atlantica T6c	3.5	3	98.50	2.81
Pseudoalteromonas haloplanktis TAC125**	3.83	5	101.68	2.65
Psychrobacter arcticus 273-4	2.65	7	132.28	4.99
Psychrobacter cryohalolentis K5	3.06	5	193.26	6.32
Psychrobacter sp. PRwf-1	2.98	10	204.71	6.87
Shewanella baltica OS155	5.12	22	323.04	6.31
Shewanella baltica sp. MR-4	4.7	9	177.94	3.79
Shewanella baltica sp. MR-7	4.07	9	162.43	3.99
Shewanella denitrificans OS217	4.54	7	110.83	2.44
Vibrio cholerae O395**	4.12	10	265.41	6.44

Table 2SM. *cont.*

Alphaproteobacteria (16)				
Brevundimonas sp. BAL3	3.63	6	114.08	3.14
Candidatus *Pelagibacter ubique* HTCC1062	1.3	1	13.36	1.03
Jannaschia sp. CCS1	4.31	7	147.01	3.41
Nitrobacter winogradskyi Nb-255	3.4	13	234.24	6.89
Pelagibacter sp. HTCC7211	1.4	0	0	0
Rhodobacter sphaeroides ATCC17025	3.21	14	389.03	12.12
Rhodobacter sphaeroides KD131**	4.34	9	130.48	3.01
Roseobacter denitrificans Och 114	4.1	9	221.78	5.41
Roseobacter sp. CCS2	3.49	4	63.83	1.83
Roseobacter sp. MED193	4.65	6	83.90	1.8
Silicibacter pomeroyi DSS-3	4.1	7	194.32	4.74
Silicibacter sp. MT1040	3.2	1	29.64	0.93
Silicibacter sp. TrichCH4B	4.67	6	109.43	2.34
Sphingomonas wittichii RW1	5.38	8	181.36	3.37
*Sulfitobacte*r sp. EE36	3.54	4	114.15	3.22
Thalassiobium sp. R2A62	3.48	10	185.42	5.33
Bacteroidetes (21) (14 Flavobacteria + 7 Non Flavobacteria)				
Bacteroides fragilis NCTC 9343	5.2	13	253.27	4.87
Bacteroides fracilis YCH46	5.3	15	274.44	5.18
Bacteroides thetaiothaomicron VPI-5482	6.2	22	340.91	5.50
Bacteroides vulgatus ATCC 8482	5.1	22	436.62	8.56
Porphyromonas gingivalis ATCC 33277	2.3	7	124.94	5.43
Salinibacter ruber DSM 13855	3.5	6	154.4	4.41
Salinibacter ruber M8	3.6	7	123.54	3.43
Croceibacter atlanticus HTCC2559	2.9	1	41.81	1.44
Cytophaga huchinsonii ATCC33406	4.4	4	136.96	3.11
Dokdonia sp. MED134	3.3	2	38.98	1.18
Flavobacteria bacterium BBFL7	3.1	0	0	0.00
Flavobacteriales bacterium ALC-1	3.8	0	0	0.00
Flavobacterium bacterium BAL38	2.8	3	49.464	1.77
Flavobacterium johnsoniae UW101	6.1	9	247.25	4.05
Flavobacterium psychrophilum JIP02/86	2.8	0	0	0.00
Gramella forsetii KT0803	3.8	5	105.29	2.77
Kordia algicida OT-1	5	7	173.13	3.46
Leeuwenhoekiella blandensis MED217	4.2	3	54.80	1.30
Polaribacter irgensii 23-P	2.7	1	14.90	0.55
Polaribacter sp. MED152	2.9	1	12.51	0.43
Robiginitalea biformata HTCC2501	3.5	6	157.84	4.51

* RATIO: Genomic islands size (kb) x100/Genome size (kb).

** Bacterial genomes with two chromosomes in which we have integrated all genomic islands found in both chromosomes.

[a] We shown only those GIs ≥ 9.5 kb.

Table 3SM. Taxa specific gene enrichment analyses

Bacterial Groups Comparison*	# Genes in Reference Set	# Genes in Test Set	# Annotated Genes (Ref/Test)	Fisher´s Significance Test**
All-A (Ref) vs. A (Test)	6211	2117	2523/837	YES
All-C (Ref) vs. C (Test)	7246	1082	3026/334	YES
All-D (Ref) vs. D (Test)	6863	1465	2894/466	YES
All-E (Ref) vs. E (Test)	7166	986	2875/404	YES
All-G (Ref) vs. G (Test)	5826	2502	2122/1238	YES

Fisher´s exact Test (FT) was used to detect bacterial taxa enriched genes within GIs by comparison between different combinations of bacterial groups. We referred as Reference Set (Ref) all genes within GIs for all phylogenetic groups except with the
taxa to compare with (Test Set).

*The five bacterial group tested are labeled as follow: A=Alphaproteobacteria; C=Flavobacteria; D= Non marine Bacteroidets; E= Cyanobacteria, G=Gammaproteobaacteria.
** Positive results of the Fisher´s exact Test. These categories with statistical significance (p-value <0.01) are listed in Table 2.

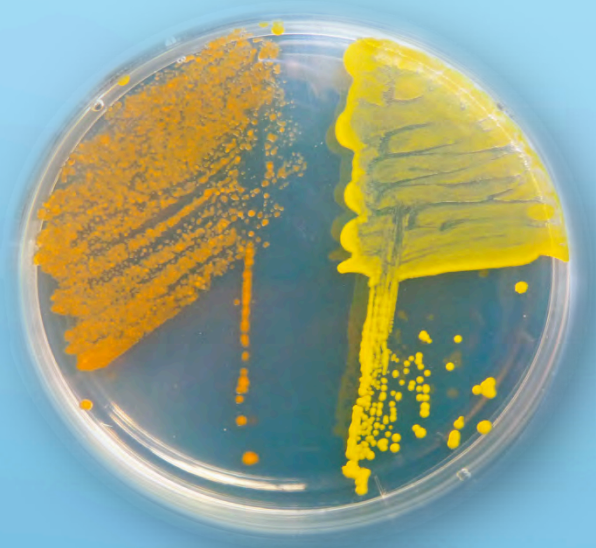
Table 4SM. Matrix showing the number of genes within GIs assigned to each of the 16 biological categories for all marine bacterial genomes analyzed.

Bacterial strain[a]	Photosynthesis[b]	Energy Metabolism Enzymes	Ribosomal proteins	Hydrolysis	Polysacch. Biosynthesis	DNA-Restrict. Modif. System	DNA-directed RNA pol.	Transport	Two Comp. System	Cell Motility	Stress response	MGE	CRISPRs	Virulence	TA-toxins	Secretion System
ATCC 29413	3	0	4	4	1	5	0	1	3	0	1	3	4	1	2	0
PCC 7424	0	0	1	4	0	0	0	0	0	1	0	0	0	1	0	0
PCC 8801	0	1	0	1	0	3	1	0	0	0	0	3	0	0	1	0
PCC 73102	1	18	1	4	3	0	1	0	4	0	5	2	0	2	0	0
PCC 7120	12	1	0	2	2	0	0	1	2	0	0	0	0	0	0	0
AS9601	0	2	0	0	0	1	0	0	0	0	0	1	0	0	0	0
MIT 9215	0	2	0	0	0	1	0	0	0	0	0	1	0	0	0	0
MIT9312	2	2	0	2	1	1	0	2	0	0	0	1	0	0	0	0
NATL1A	0	6	0	0	2	0	1	0	0	0	0	0	0	0	0	0
NATL2A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PCC 6301	0	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0
CC9311	1	7	0	1	1	0	0	2	0	1	0	1	0	0	0	0
CC9605	7	2	2	4	0	0	2	0	1	0	0	13	0	2	0	0
RCC307	0	5	0	0	0	0	0	1	0	0	0	0	0	0	0	0
WH7803	0	1	1	0	0	0	0	8	0	0	0	0	0	0	0	1
PCC6803	3	5	0	1	2	0	1	8	3	0	0	0	0	0	0	0
"D. ecotype"	0	1	0	4	1	4	1	12	4	8	0	10	0	0	1	0
DSM 3043	0	0	0	1	0	1	1	3	0	0	0	2	0	0	0	0
34H	0	0	0	3	0	1	0	0	0	0	0	2	0	0	0	0
L2TR	0	1	0	8	1	0	0	5	1	0	1	7	0	0	0	1
VT8	0	0	0	6	4	4	0	6	1	1	2	12	0	1	2	1
MWYL1	0	1	0	1	1	1	0	0	1	0	0	4	0	0	2	0
ATCC 19707	0	0	0	1	1	4	2	0	2	0	1	6	0	2	3	1
T6c	0	0	0	3	0	1	0	2	1	0	0	27	0	0	0	0
TAC125	0	2	0	0	0	1	2	5	2	2	1	4	0	0	0	0
273-4	0	2	0	4	2	1	0	2	0	0	0	11	0	0	0	0
K5	0	4	0	0	2	11	0	2	0	1	0	8	0	0	0	0
Prwf-1	0	3	0	0	6	3	0	8	2	0	0	13	0	0	0	0
OS155	0	0	26	3	12	1	3	3	4	9	1	31	0	0	0	0

Table 4SM. *cont.*

OS217	0	4	0	2	2	1	1	1	0	1	0	3	0	0	0	0
MR-4	0	2	0	1	5	3	1	5	0	0	2	11	0	0	0	0
MR-7	0	2	0	2	2	2	0	2	1	0	0	5	0	0	0	0
O395	0	0	29	1	2	0	1	5	0	1	0	9	0	19	8	1
BAL3	0	5	0	1	0	0	0	6	0	0	2	5	0	0	1	1
HTCC1062	0	0	0	1	0	0	0	2	0	0	0	1	0	0	0	0
CCS1	0	0	0	0	0	0	0	2	4	0	0	11	0	0	0	1
Nb-255	0	5	0	1	1	4	0	3	1	0	3	30	0	0	1	0
ATCC17025	0	5	0	2	0	0	2	9	3	3	0	33	0	0	1	0
KD131	0	1	3	2	1	0	0	5	0	0	0	2	7	0	0	0
Och 114	0	0	0	1	0	10	1	5	3	21	0	8	0	0	0	0
CCS2	0	3	1	0	1	0	0	5	1	0	0	5	0	0	0	0
MED193	0	4	0	3	0	3	0	0	0	11	0	4	0	0	0	0
DSS-3	0	4	18	5	0	8	2	0	0	0	0	8	0	0	0	1
TM1040	0	1	0	1	0	0	0	0	0	0	0	12	0	0	0	0
TrichCH4B	0	0	0	2	1	2	2	5	2	0	1	3	0	0	0	1
RW1	0	0	5	4	2	0	2	4	0	0	0	12	0	0	0	0
EE36	0	0	32	3	0	1	4	1	0	0	0	7	0	0	0	0
R2A62	0	3	3	3	1	2	0	4	2	0	1	10	0	0	0	7
BAL3	0	5	0	1	0	0	0	6	0	0	2	5	0	0	1	1
M8	0	2	0	0	1	0	2	2	0	1	0	6	0	0	2	0
HTCC2559	0	0	0	1	0	1	0	3	0	0	1	1	0	0	0	0
ATCC 33406	0	0	0	3	0	5	0	9	0	1	0	2	0	0	0	0
MED134	0	0	0	0	0	0	0	5	0	0	1	2	0	0	0	0
BAL38	0	0	5	0	0	0	2	0	0	0	0	0	0	0	0	0
UW101	0	0	5	7	0	0	5	3	2	0	1	23	0	2	0	0
KT0803	0	3	5	2	0	0	2	2	0	0	0	0	0	0	3	0
OT-1	0	0	5	0	0	5	2	0	1	0	1	1	0	0	3	0
MED217	0	0	0	0	0	0	0	1	0	0	0	3	4	0	0	0
23-P	0	1	4	0	0	0	2	0	0	0	0	0	0	0	0	0
MED152	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

[a] The complete bacterial name is shown in Table 3SM in the same order. Here only the name of the strain is shown.
[b] Each of the 16 biological categories are detailed in Material & Methods and placed in the same order as in Figure 10.



Chapter 4

Culture medium influences the response to light of PR-containing Flavobacteria

Beatriz Fernández-Gómez, José M. González and Carlos Pedrós-Alió

ABSTRACT

Proteorhodpsin (PR) is a photoactive protein present in marine bacterioplankton, including Bacteria and Archaea, and Eukarya. It consists of a transmembrane protein bound to a retinal molecule and functions as a light-driven proton pump. Although it is very abundant in the ocean, its physiological function is still not well understood, since it has been suggested that it could play a different role in each of the studied strains that carry this gene. Recent studies have demonstrated that light enhances cell yield and carbon fixation in two different PR-carrying bacteria (*Dokdonia* sp. MED134 and *Polaribacter* sp. MED152, respectively), suggesting different strategies in use of the PR. Here we demonstrate that light stimulates cell yield under low carbon growth conditions also in *Polaribacter* sp. MED152. We also confirm that this enhancement depends not just on the concentration but also on the nature of the labile organic matter available since it does not occur when, under the same incubation conditions, the carbon source changes.

INTRODUCTION

Proteorhodopsins (PRs) were originally discovered in a fosmid clone from a metagenomic library (Béjà *et al*., 2000). By cloning of the gene in *E. coli*, the protein was shown to be a retinal-binding transmembrane protein that worked as a light-driven proton pump (Béjà *et al*., 2000; Béjà *et al*., 2001; Wang *et al*., 2003). In the few years after its discovery, PRs were found to be present in phylogenetically diverse groups of marine bacteria (de la Torre *et al*., 2003; Frigaard *et al*., 2006). Moreover, metagenomic surveys found PRs to be very abundant in the surface ocean (Rusch *et al*., 2007; Venter *et al*., 2004) to the point that the relative abundance of PR-containing bacteria from oceanic surface waters has been estimated to be between 13% and 80% (de la Torre *et al*., 2003; Rusch *et al*., 2007; Sabehi *et al*., 2005; Venter *et al*., 2004). There are also indications that PR is commonly acquired and maintained by lateral gene transfer (Frigaard *et al*., 2006; McCarren and DeLong, 2007; Sharma *et al*., 2006), suggesting its possession confers certain ecological advantages. Thus, any bacterium that contains the genes to synthesize ß-carotene can become a phototroph by just incorporating two genes: the gene coding for the opsin protein (*prd*), and the gene coding for the enzyme that converts ß-carotene into retinal (*blh*). It is not surprising, therefore, that PRs are so broadly distributed and ubiquitous in bacteria inhabiting the ocean's photic zone.

Despite the alleged role of PRs in energy conservation, a clear ecological function has not been demonstrated (DeLong and Béjà, 2010; Fuhrman *et al*., 2008; Lami *et al*., 2009). In part, this is due to the fact that there were very few PR-containing bacteria isolated in pure culture. Initial studies with the Proteobacteria *Pelagibacter ubique* and isolates from the SAR92 clade were frustrating (Giovannoni *et al*., 2005; Stingl *et al*., 2007). Neither one of these bacteria showed any differences in growth between light and dark incubations under different culture conditions. Recently, *Vibrio* AND4 was shown to survive better in the light than in the dark thanks to PR (Gómez-Consarnau *et al*., 2010) and *P. ubique* has been shown to reduce respiration rates in the light with respect to the dark, in periods of carbon starvation, apparently obtaining ATP through the PR (Steindler *et al*., 2011). It seems, therefore, that one of the roles of PR, at least in Proteobacteria, is to help the bacteria survive under carbon starvation conditions.

Gómez-Consarnau *et al.* (2007) showed that the Flavobacteria *Dokdonia* sp. MED134 produced higher yields in the light than in the dark at very low organic matter concentrations. *Dokdonia* sp. MED134 did not grow when organic matter concentration was minimal, and it grew to similar growth yields in the light and in the dark when concentrations of organic matter were those typically high of Zobell medium. However, at organic matter concentrations similar to

those usually found in the surface ocean, growth yields were up to four times higher in the light. A similar response could not be demonstrated in *Polaribacter* sp. MED152, another marine Flavobacteria with PR. This strain showed an erratic growth behavior, sometimes forming clumps of cells and other times growing without pigments, and it was not possible to obtain clear-cut results. However, this strain showed increased CO_2 fixation in the light with respect to the dark, under the same conditions (Gónzalez *et al*., 2008).

It was unsatisfactory that these two Flavobacteria had many similarities in their genomes (Fernández-Gómez *et al*., submitted; González *et al*., 2011), and yet showed different and, as mentioned, often erratic, responses to light. In the present work, we made an attempt to systematize the growth conditions leading to better growth and/or to larger CO_2 fixation in the light in these two Flavobacteria.

MATERIALS AND METHODS

Isolation of Flavobacteria strains

Polaribacter sp. MED152 and *Dokdonia* sp. MED134 strains were isolated on 20 March 2001 from Northwest Mediterranean Sea surface water (0.5 m depth) collected 1 km off the coast at the Blanes Bay Microbial Observatory (BBMO), Spain (41º 40' N, 2º 48' E). MED152 was isolated on Marine Agar 2216 (Difco) and MED134 on Zobell agar plates (Fig. 1).

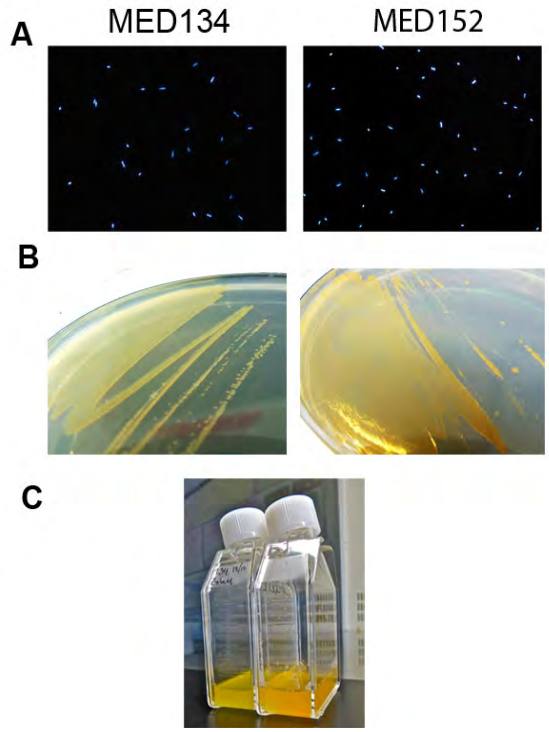


Figure 1. Images of Flavobacteria isolates. (A) Epifluorescence microscopy images of MED152 and MED134 cells stained with DAPI. **(B)** Morphology and color of colonies on Zobell agar. **(C)** Color of MED152 and MED134 growing in Zobell liquid medium.

Preparation of culture media

1. Artificial seawater. Artificial seawater was prepared with Sea Salts (Sigma®). Salts are highly hygroscopic and the weight of the salts increases due to water content with time. Thus, final salinity was checked with a salinometer (ATAGO Co. LTD) and adjusted to a salinity of 35. Seawater medium was sterilized by autoclaving (120 °C, 30 min). After cooling, pH was adjusted to 7.6 with 1N NaOH and determined with a pH meter (ORION Research, EA™ 920). Artificial seawater without any organic matter additions contained traces of dissolved organic carbon around 0.03 mM.

2. Inorganic nutrients. Nitrogen and phosphorous were provided as ammonium chloride and potassium phosphate at final concentrations of 2.1 µM NH_4Cl and 0.3 µM K_2HPO_4 µM respectively. Nitrogen and phosphorous sources were added to the artificial seawater before

sterilization. We experimented with different ways of providing iron: iron sulfate, Mohr's Salts (Ammonium iron (II) sulfate), and iron-EDTA at 100 μM. But only iron-EDTA was adequate to supply iron without forming precipitates. The iron stock solution was autoclaved separately.

3. Vitamins and other factors. Since the two bacteria lack the enzymes necessary to synthesize thiamine, biotine, and cobalanine (B12) (González *et al*., 2008), we added these vitamins at final concentrations of 0.2 mg/ml for thiamine, and 0.001 mg/ml for biotine and B12. The vitamin solution was sterilized by filtration through 0.2 μm cellulose acetate disposable syringe filters and added to the already sterile and pH balanced medium as 0.5 ml per liter of medium. After several trials, we could not get growth of the bacteria in strictly defined media. This was solved by adding a small amount of yeast extract (YE) at a final concentration of 0.05 g/L. The resulting amount of dissolved organic carbon in artificial seawater medium plus YE was 1.7 mM, including the trace carbon in artificial seawater salts. Appropriate aliquots of this solution were used to complement media with different carbon concentrations.

4. Carbon sources. Two types of carbon sources were used in different experiments. A complex medium was prepared with 5 g peptone (Bacto Peptone; BD) and 1 g yeast extract (Bacto Yeast Extract; Difco) in 1 L of artificial seawater. This is the composition of ZoBell medium. We used it at full strength for rich medium experiments and at different dilutions for tests with lower amounts of carbon sources. A second type of experiments was aimed at defining the preferred carbon sources. We assayed a combination of amino acids (combination from Sigma® at ~25 mM) or glucose (~24.5 mM) (Table 1). Carbon source solutions were sterilized by filtration through 0.2 μm cellulose acetate disposable syringe filters and added to the already sterile and pH-adjusted medium.

Table 1. Composition of the eight different media tested.

Medium number	Base	Carbon source (0.07%)	Additions
1	ASW Nitrogen, Phosphorous Vitamins	Glucose	-
2		Amino acids	-
3		Glucose	YE
4		Amino acids	YE
5		Glucose	Fe
6		Amino acids	Fe
7		Glucose	YE plus Fe
8		Amino acids	YE plus Fe

YE: yeast extract; **Fe:** iron-EDTA

Effect of light in growth and bicarbonate uptake

Incubation conditions

1. Light and temperature. Incubations were carried out in light and temperature controlled rooms. Two light regimes were used. In the first regime, cultures were incubated under continuous light. In the second regime, alternation of 12 hours light and 12 hours dark cycles was used. Dark incubations were carried out in the same incubators with aluminum foil covered bottles. Light was provided by Sylvania GRO-LUX F30W/GRO fluorescent lamps and an average irradiation of ~150 µmol photons m^{-2} s^{-1} was used. Temperature was kept at 20°C.

2. Incubation containers and shaking. We tested a variety of containers, including polycarbonate bottles (Nalgene), polyethylene bottles (Iwaki) and glass bottles (Pyrex) (Table 2). We also tried growth with and without shaking and did not find any significant differences. Therefore, all experiments were carried out without shaking.

Table 2. Containers and rinsing procedures used in different experiments.

Experiment	Bottles	Rinsing
1. Simplified media	Polyethylene	-
2. Effect of [DOM] (Glucose + YE)	Polyethylene	Spin (3 times x 4500 rpm)
3. Effect of [DOM] (peptone + YE)	Polycarbonate	Spin (3 times x 4500 rpm)
4. Bicarbonate uptake	Polycarbonate (growth) and polyethylene for radioactivity incubations	Transfer twice in the same poor medium

3. Effect of organic matter concentration experiments. In order to carry out experiments with different concentrations of carbon sources, the cultures were initially started from a colony on a Petri dish with solid rich medium. Once growth was visible, an aliquot was inoculated in a medium of intermediate organic carbon concentration (~24.5 mM C) and allowed to grow for 72 h. Then 50 ml aliquots were spun down at 4500 rpm during 10 minutes (Sigma 3-16K, rotor no. 11180) and re-suspended in artificial seawater three times to remove all the organic matter. The abundance of cells in the final concentrate was determined by epifluorescence microscopy and cells were added to fresh medium with the desired concentration of organic matter to reach an initial abundance of 10^5 cells ml^{-1}.

Growth measurements

For the first experiment, bacterial growth was determined by following absorbance at 650 nm. For the remaining experiments, samples were fixed with glutaraldehyde (10% final concentration), stained with DAPI (Porter and Feig, 1980), filtered on black 0.2 µm-pore-size polycarbonate filters and counted by epifluorescence microscopy within 48 hours.

Bicarbonate uptake

Bacteria were first grown in an intermediate medium (~24.5 mM C) at 20°C. When the cultures reached high biomass, 10 ml were transferred to a poor medium (0.1 mM C, glucose + yeast extract) and the bottles were divided into light (~150 µmol photons m^{-2} s^{-1}) and dark conditions (bottles covered with aluminum foil and black plastic bags). After 22h in the poor culture (at the end of the exponential phase), 10^{-5} cells ml^{-1} were transferred again in to the same poor medium and incubated under light or dark conditions. After 22 hours, five subsamples from each culture were transferred to 25 cm^3 polyethylene tissue flasks (50 ml, Iwaki). Three were treated as experimental and two as controls. All bottles were incubated under white light (no light entered the dark bottles) for two hours with 0.5 ml of $H^{14}CO_3^-$ (20µCi; DHI). Controls were fixed with glutaraldehyde at 10% final concentration. After incubation, the 50 ml were filtered through GSWP 0.22 µm pore size filters (Millipore) and the filters were placed in scintillation vials and exposed to HCl (10 M) fumes overnight. Finally, 4 ml of scintillation mixture (Perkin Elmer) were added and kept 24h before counting with a liquid scintillation analyzer (Perkin Elmer). Bicarbonate uptake rates were calculated from the equation given in Vollenweider (1974). At the time of the experiment bacterial abundance was ~1.5×10^6 cells ml^{-1} on average for both bacteria, as determined by epifluorescence microscopy of DAPI stained preparations.

Determination of dissolved organic carbon (DOC)

DOC present in the medium was measured with a Shimadzu TOC-V organic carbon analyzer. Approximately 10 ml of medium were collected in precombusted glass ampoules, adding 20µl of H_3PO_4 to fix the samples and the ampoules were heat-sealed and stored at 4ºC until analysis. Each ampoule was injected three to five times, yielding a standard deviation of <1% to determine the concentration of DOC.

The performance of the analyzer was tested with the DOM reference material provided by Prof. D. Hansell from the University of Miami. We obtained a concentration of 45.2 ± 0.3 µmol C L^{-1} for the deep ocean reference (Sargasso Sea deep water, 2600 m) minus blank reference

materials, the day when samples were analyzed. The nominal DOC value provided by the reference laboratory is 45 µmol C L^{-1}.

RESULTS

1. Growth of Polaribacter and Dokdonia in simplified media

Different combinations of carbon sources and other nutrients were tested in order to determine the best combination (Table 2) for both bacteria to grow, always trying to use the minimal possible number of components.

There was a large difference in growth with and without additions of YE in all cases (Fig. 2, filled versus empty symbols). Both bacteria showed higher growth rates and yields and

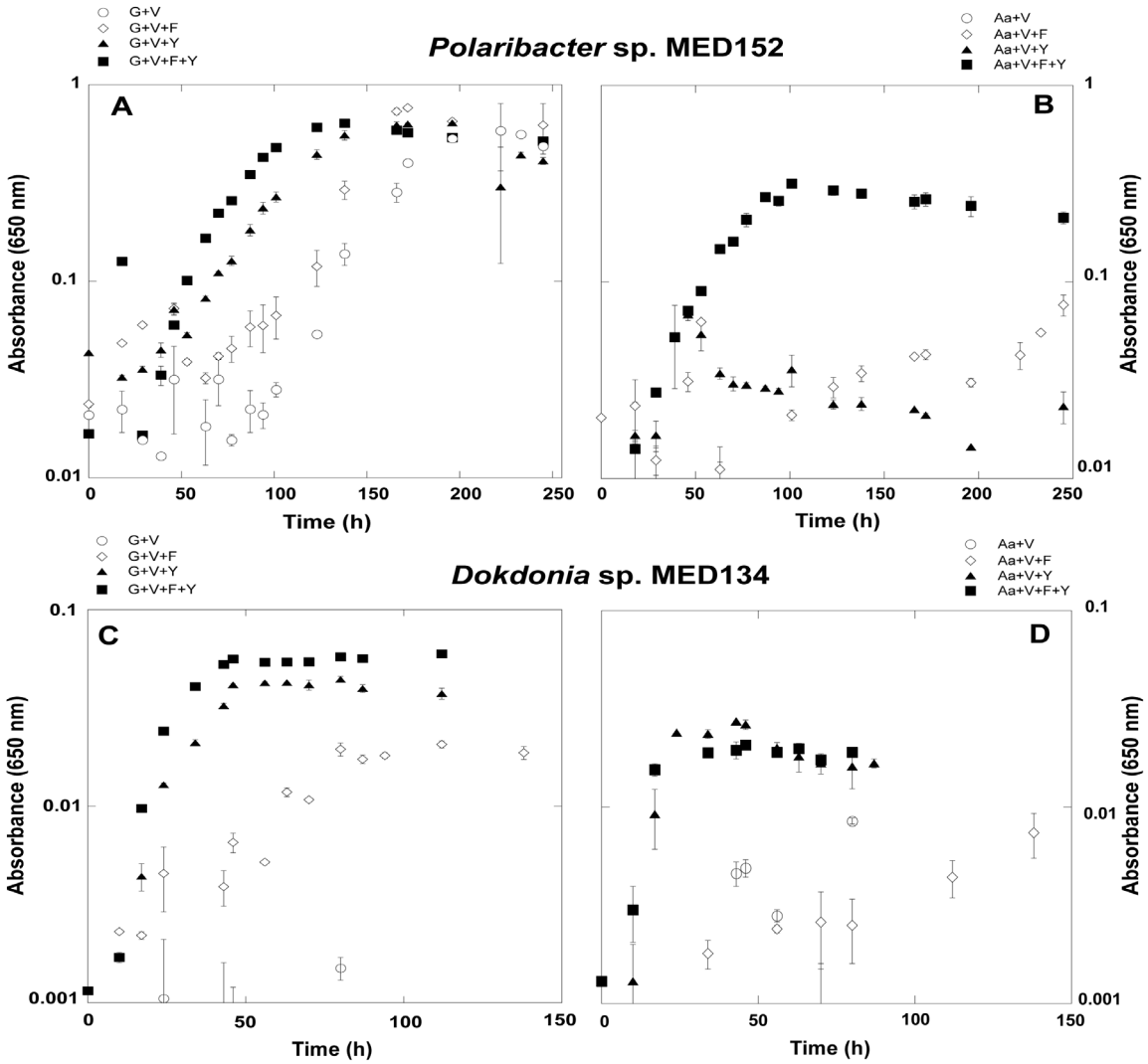


Figure 2. Growth curves for both Flavobacteria in eigth simplified media. Error bars show range of values for duplicate cultures.

shorter lag phases when YE was present (Fig. 3 and Table 3). In almost all cases *Polaribacter* had longer lag phases and slightly lower growth rates than *Dokdonia*. When YE was present in the medium, Fe caused very minor variations. Only in the case of MED152 growing on amino acids the absence of iron had a significant effect on the yield (Fig. 2B).

In the absence of YE, growth with glucose was slower and yields were lower for both bacteria (Fig. 2A and C, empty versus filled symbols). Presence or absence of iron did not make a significant difference. When amino acids were the main carbon source, however, both bacteria showed either very slow growth rates (*Dokdonia*, Fig. 2D) or erratic and minimal In the absence of YE, growth with glucose was slower and yields were lower for both bacteria (Figures 2A and C, empty versus filled symbols). Presence or absence of iron did not make a significant difference. When amino acids were the main carbon source, however, both bacteria showed either very slow growth rates (*Dokdonia*, Fig.2D) or erratic and minimal growth (*Polaribacter*, Fig. 2B).

Combining the results from these experiments, a medium with glucose and YE was obviously the medium of choice.

2. Growth of Polaribacter and Dokdonia at different organic matter concentrations

A series of organic carbon concentrations (DOM, as yeast extract plus glucose from 0.03 mM C to 250 mM carbon) was prepared for the growth of MED152 and MED134, both in the light and in the dark (Fig. 3). Essentially no growth was observed in the unenriched cultures containing only the trace DOM in the sea salts (Fig. 3, Fig. 4, Table SM1). Growth rates increased with DOM additions until 0.2 mM carbon, when maximal growth rates of

Table 3. Growth rate (μ), lag and yield for each of the eight combinations of simplified media.

		MED152			**MED134**		
Medium #	Additions	μ (h^{-1})	Lag (h)	Yield	μ (h^{-1})	Lag (h)	Yield
1	Glu	0.03	94	0.01	0	-	0
2	Aa	0	-	0	0	-	-
3	Glu +YE	0.03	101	0.7	0.05	-	0.01
4	Aa +YE	0	-	0	0.02	87	-
5	Glu+ Fe	0.03	53	0.58	0.05	24	0.04
6	Aa +Fe	0	-	0	0.17	10	0.02
7	Glu +Fe +YE	0.07	29	0.59	0.02	10	0.05
8	Aa +Fe+ YE	0.05	-	0.31	0.13	10	0.02

Glu: glucose; **Fe:** iron-EDTA; **YE:** yeast extract; **Aa:** amino acids

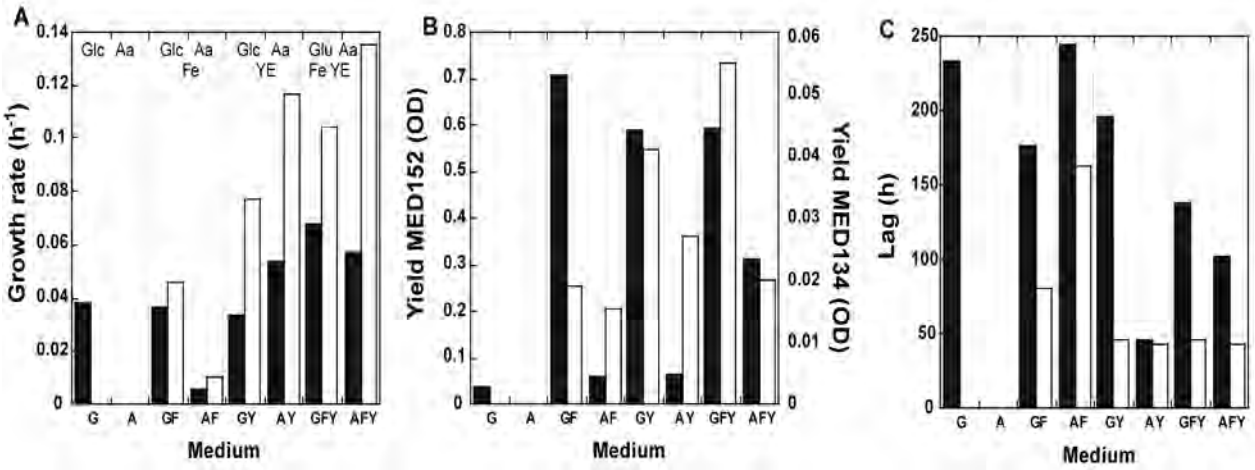


Figure 3. Summary of growth rates, yields, and hours to reach maximal yield in light and dark incubations in different simplified media, with glucose or aminoacids as carbon sources. Black bars: MED152, white bars: MED134.

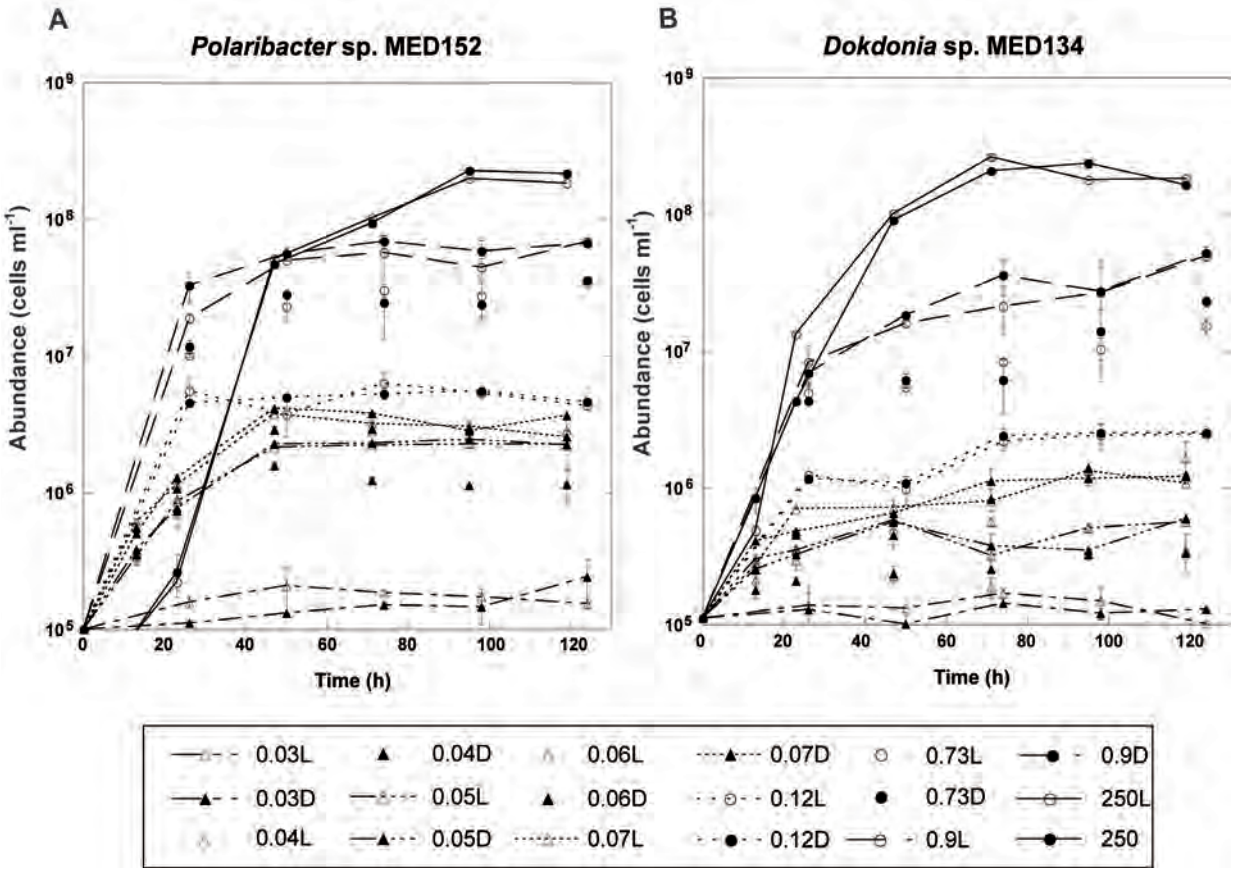


Figure 4. Cultures grown in artificial seawater enriched with different concentrations of organic matter (glucose + yeast extract) exposed to light (open symbols) and dark (filled symbols) conditions. Error bars show range of values for duplicate cultures.

approximately 0.1 day^{-1} were reached (Fig. 5A). Values were very similar for both bacteria. In the case of *Polaribacter*, the 250 mM carbon of ZoBell medium actually caused a slight decrease in the growth rate. Yields increased very steeply until 0.2 mM carbon and then kept

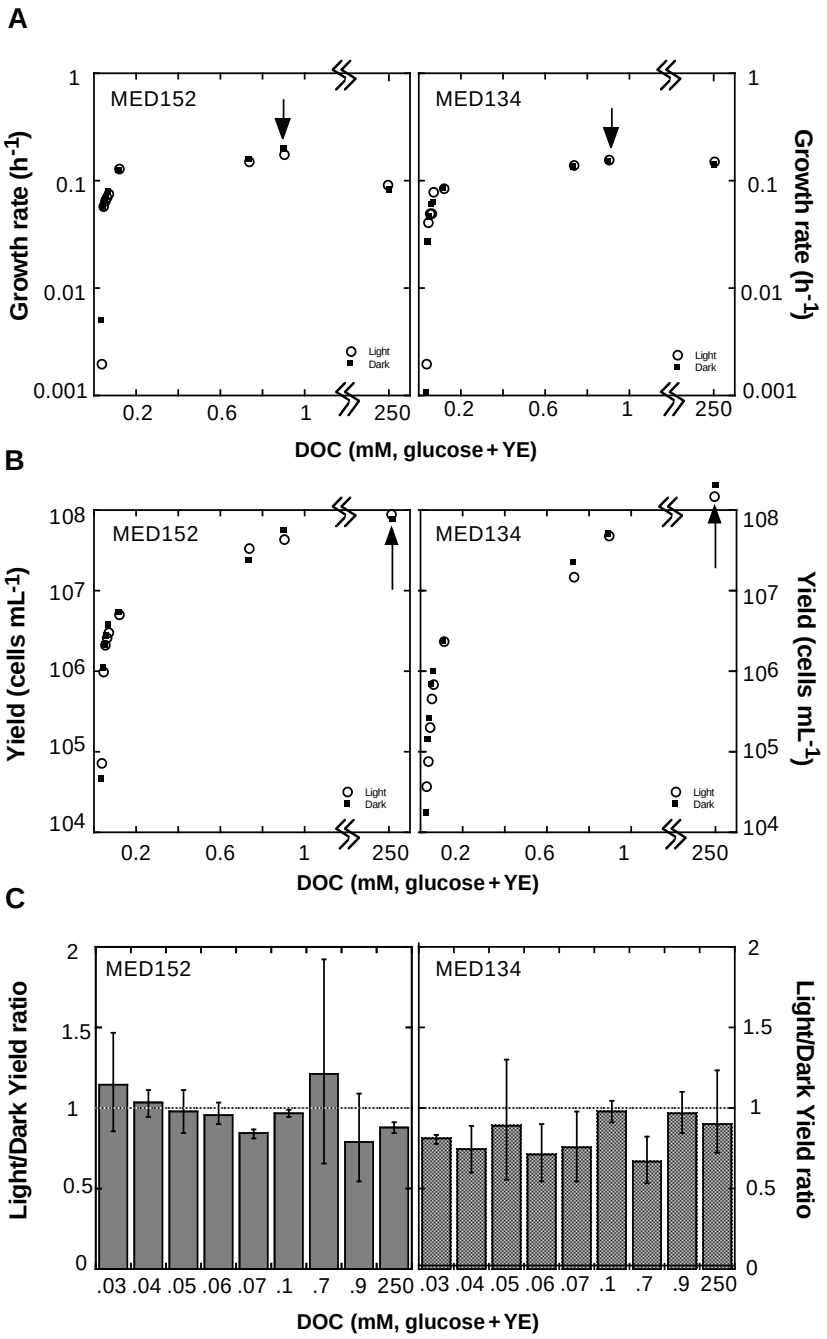


Figure 5. Growth rate (μ) and yield versus organic matter concentration (glucose + yeast extract), and proportion between yield in the light (open symbols) and in the dark (filled symbols) .

increasing at a slower pace with increasing DOC concentrations (Fig. 5B).

There were no significant differences in growth rates between light and dark incubations. When yields were compared (Figure 5C) there were no significant differences in *Polaribacter*. In *Dokdonia* there were some cases of actually lower yield in the light than in the dark, but the differences were always very small.

3. Growth of MED152 in complex medium

Gómez-Consarnau *et al*. (2007) showed larger yields in the light than in the dark for *Dokdonia* growing on peptone and YE. The same conditions were used to grow *Polaribacter*. In artificial seawater without enrichment (0.03 mM C) little growth was observed in light or darkness. Enrichment with DOM corresponding to 0.11 mM C (final concentration) resulted in bacterial cell yields about twofold higher in the light than in the dark (Fig. 6). Greater carbon enrichment (0.37 or 0.72 mM C) gave yields about 1.6-fold higher in the light cultures. At higher concentrations (0.91 mM C) the differences started to be smaller, giving a yield of around 1.2-fold higher in light incubations. In the full-strength medium (242 mM C), similar bacterial cells yields were observed in both light and dark conditions. No significant

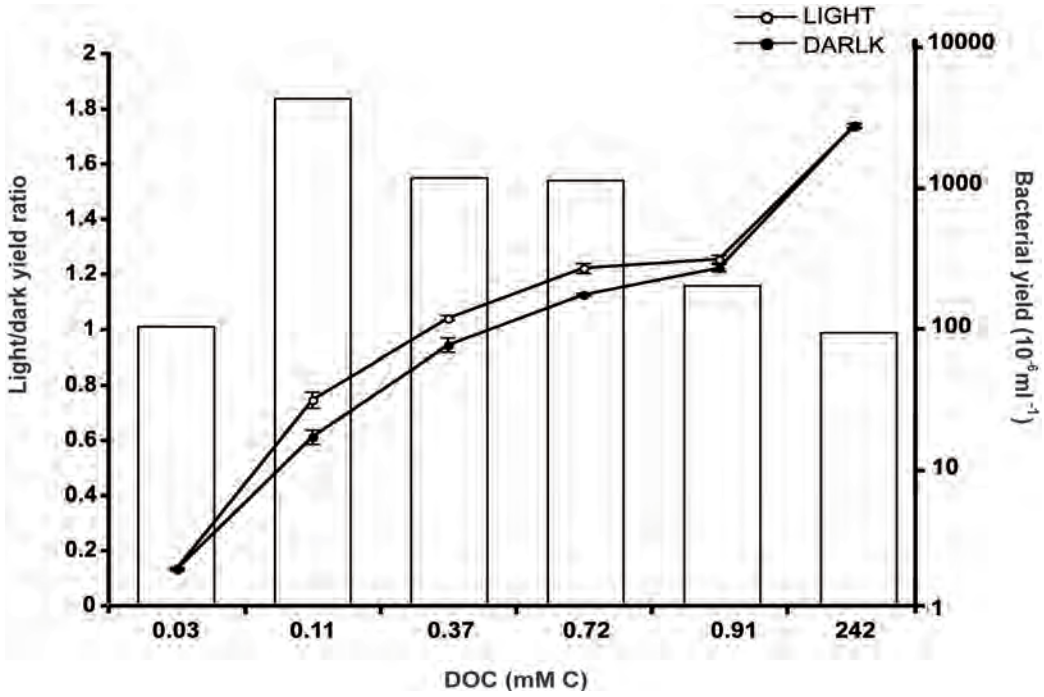


Figure 6. Growth of MED152 in seawater cultures with different concentrations of DOM (peptone + yeast extract). Cell yields in artificial seawater, incubated in the light (open circles) and in the dark (filled circles), enriched with dissolved organic matter, in comparison with unenriched controls (0.03 mM C) and full-strength medium (242 mM C). Columns represent the light/dark ratio in cell yield. Error bars denote standard deviation for triplicate cultures.

differences in cell volume were observed either among the cultures grown in the light and the ones grown in the dark or among the different concentrations (an average of 0.09 μm^3).

4.Differences in bicarbonate uptake under light and dark conditions.

Bicarbonate incorporation by MED152 and MED134 in the light and dark were studied. The medium used was the glucose plus YE medium developed before. Although no differences in growth between light and dark incubations were observed in any of the two organisms, MED152 showed a significant difference in bicarbonate incorporation between light and dark incubations: there was 3.6-fold more fixation in the light than in the dark. Bicarbonate uptake in MED134 was practically the same for both conditions showing no significant differences between light and dark (Fig. 7A).

No significant differences in cell volume were observed in any of the two strains. MED152 had an average volume of 0.07 μm^3 both under light and dark conditions. MED134 cells had an average volume of 0.126 μm^3 and 0.110 μm^3 in light and dark conditions respectively.
When bacteria grew in the same conditions but in a light-dark cycle (instead of 100% light), and in a richer medium (~24.5 mM C, instead of 0.1 mM C), the differences in bicarbonate uptake were 2-fold higher fixation in the light for MED152 while MED134 showed an increase of 1.6 times more fixation in the light (Fig. 7B).

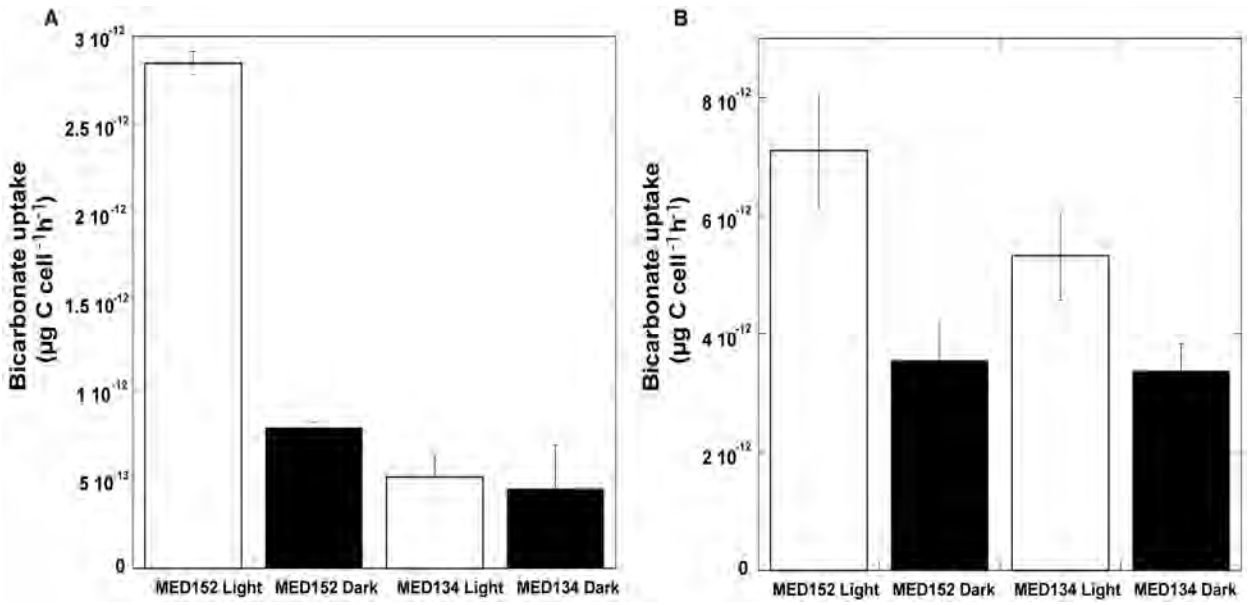


Figure 7. (A) Bicarbonate uptake in MED152 and MED134 under continuous light (white bars) and in the dark (black bars) (0.1 mM C). **(B)** Bicarbonate uptake in MED152 and MED134 under a light-dark cycle and in the dark (24.5 mM C). Error bars show standard deviations of triplicates.

DISCUSSION

Dokdonia MED134 was the first PR-containing bacterium in which a differential response to light could be demonstrated in culture (Gómez-Consarnau *et al*., 2007). These results were confirmed by Kimura *et al*. (2011), who also analyzed the transcriptome of this bacterium. When the genomes of *Dokdonia* MED134 and *Polaribacter* MED152 were annotated (González *et al*., 2008, González *et al*., 2011) the two bacteria had a similar suite of genes related to PR, pigments, light sensing, anaplerotic carbon fixation and polymer degradation. Thus, it was frustrating that the effect of light on *Polaribacter* MED152 could not be demonstrated due to the erratic growth response of this bacterium. In the present work we were successful in finding culture conditions that would reproductively lead to growth of the bacterium. When the experiment of Gómez-Consarnau *et al*. (2007) was repeated with *Polaribacter* MED152 we could show a difference in yield between light and dark incubations. This difference was slightly smaller than that found in *Dokdonia*, but it was nevertheless significant. This makes MED152 the second bacterium showing increased yields in the light thanks to the presence of PR.

On the other hand, an increased anaplerotic bicarbonate fixation in the light had been found in *Polaribacter*, but no clear response had been seen in *Dokdonia*. We were successful in finding conditions under which *Dokdonia* also showed this response. Moreover, the light stimulation was not significant under continuous light. Only when a 12 h light/12 h dark cycle was used did we find a significant response. It is intriguing that the response to light of *Dokdonia* was larger in terms of cell yield but smaller in terms of bicarbonate fixation. These differences suggest a fine tuning so that the two bacteria adapt to slightly different niches. These variations within a common theme were already noted in a comparison of the genomes of these bacteria (Fernández-Gómez *et al*., submitted). They had very similar number of proteases and glycoside hydrolases, for example, but the particular types present in each genome were slightly different.

Another interesting observation concerned the type of medium used. Gómez-Consarnau *et al*. (2007) had shown that the differential response to light only occurred at low (but above background) organic matter concentrations. Here we found that the differential response was not only dependent on the concentration of organic matter but also on its quality. None of the many experiments carried out in the simplified medium resulted in significant differences between growth in the light or in the dark in either one of the two bacteria. The main difference between the diluted ZoBell and the simplified medium was the presence of peptone in the former and glucose in the latter. Peptone is a product of the

hydrolysis of proteins and includes a mixture of peptides and free amino acids. We showed here that, under the conditions assayed, amino acids were a worse substrate than glucose for both bacteria. However, examination of their genomes showed the presence of a very large number of proteases per megabite of genome. Moreover, the number of proteases was much higher than that of glycoside hydrolases and this was the case for all the marine Flavobacteria with sequences genomes. In situ microautoradiography studies have also shown a preference of marine Flavobacteria for proteins (Cottrell and Kirchman, 2000). Considering all these pieces of evidence together we had expected amino acids to be a better food source than glucose. Obviously this was not the case. Apparently, proteins and peptides (polymers) are the preferred carbon source, but glucose is preferable to amino acids (monomers). The experiment of growing the bacteria on solid surfaces with proteins as carbon source under light and dark conditions seems a good possibility to clarify this issue.

In summary, there are two traits of PR-containing bacteria that begin to emerge from different studies. The first one is that each bacterium uses PR in a slightly different way from its close relatives. And the second is that demonstrating the effects of PR is going to require a substantial amount of trial and error experiments with different culture media and growth conditions.

ACKNOWLEDGEMENTS

B. F.-G. was recipient of a I3P grant from CSIC. We thank J. Pinhassi and L. Gómez-Consarnau for their advice on growing MED152 and MED134, Mara Abad for her help with dissolved organic carbon analysis, and Claudia Egas for her help with cell counts. This work was supported by grant “GEMMA” CTM2007-63753-C02-01/MAR from the Spanish Ministerio de Ciencia e Innovación.

REFERENCES

Béjà O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP *et al.,* (2000). Bacterial Rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289:** 1902-1906.

Béjà O, Spudich EN, Spudich JL, Leclerc M, Delong EF. (2001). Proteorhodopsin phototrophy in the ocean. *Nature* **411:** 786-789.

Cottrell MT, Kirchman DL. (2000). Natural assemblages of marine proteobacteria and members of the Cytophaga-Flavobactera cluster consuming low- and ligh-molecular-weight dissolved organic matter. *Appl Environ Microbiol* **66:** 1692-1697.

de la Torre JR, Christianson LM, Béjà O, Suzuki MT, Karl DM, Heidelberg J *et al.,* (2003). Proteorhodopsin genes are distributed among divergent marine bacterial taxa. *Proc Natl Acad Sci U S A* **100:** 12830-12835.

DeLong EF, Béjà O. (2010). The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times. *PLoS Biol* **8:** e1000359.

Fernández-Gómez B, Richter M, Schüler M, Pinhassi J, Acinas SG, González JM *et al.,* (2011). Comparative genomics of marine members of Bacteroidetes with and without proteorhodoposin. *The ISMEJ (Submitted).*

Frigaard N-U, Martínez A, Mincer TJ, DeLong EF. (2006). Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature* **439:** 847-850.

Fuhrman JA, Schwalbach MS, Stingl U. (2008). Proteorhodopsins: an array of physiological roles? *Nat Rev Micro* **6:** 488-494.

Giovannoni SJ, Bibbs L, Cho J-C, Stapels MD, Desiderio R, Vergin KL *et al.,* (2005). Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438:** 82-85.

Gómez-Consarnau L, González JM, Coll-Lladó M, Gourdon P, Pascher T, Neutze R *et al.,* (2007). Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. *Nature* **445:** 210-213.

Gómez-Consarnau L, Akram N, Lindell K, Pedersen A, Neutze R, Milton DL *et al.,* (2010). Proteorhodopsin phototrophy Promotes Survival of Marine Bacteria during Starvation. *PLoS Biol* **8:** e1000358.

González JM, Pinhassi J, Fernández-Gómez B, Coll-Lladó M, González-Velázquez M, Puigbó P *et al.,* (2011). Genomics of the proteorhodopsin-containing marine flavobacterium *Dokdonia* sp. MED134. *Appl Environ Microbiol* **77:** 8676-8687

Gónzalez JM, Fernández-Gómez B, Fernàndez-Guerra A, Gómez-Consarnau L, Sanchez O, Coll-Lladó M *et al.,* (2008). Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (Flavobacteria). *Proc Natl Acad Sci U S A* **105:** 8724-8729.

Lami R, Cottrell MT, Campbell BJ, Kirchman DL. (2009). Light-dependent growth and proteorhodopsin expression by Flavobacteria and SAR11 in experiments with Delaware coastal waters. *Environ Microbiol* **11:** 3201-3209.

McCarren J, DeLong EF. (2007). Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla. *Environ Microbiol* **9:** 846-858.

Porter KG, Feig YS. (1980). The Use of DAPI for Identifying and Counting Aquatic Microflora. *Limnol Oceanogr* **25:** 943-948.

Rusch D, Halpern A, Sutton G, Heidelberg K, Williamson S, Yooseph S *et al.,* (2007). The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* **5:** e77.

Sabehi G, Loy A, Jung K-H, Partha R, Spudich JL, Isaacson T *et al.,* (2005). New insights into metabolic properties of marine bacteria encoding proteorhodopsins. *PLoS Biol* **3:** e273.

Sharma AK, Spudich JL, Doolittle WF. (2006). Microbial rhodopsins: functional versatility and genetic mobility. *Trends in Microbiology* **14:** 463-469.

Steindler L, Schwalbach MS, Smith DP, Chan F, Giovannoni SJ. (2011). Energy starved Candidatus *Pelagibacter ubique* substitutes light-mediated ATP production for endogenous carbon respiration. *PLoS ONE* **6:** e19725.

Stingl U, Desiderio RA, Cho J-C, Vergin KL, Giovannoni SJ. (2007). The SAR92 Clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. *Appl Environ Microbiol* **73:** 2290-2296.

Venter C, Remington K, Heidelberg J, Halpern A, Rusch D, Eisen J *et al.,* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66-74.

Wang W-W, Sineshchekov OA, Spudich EN, Spudich JL .(2003). Spectroscopic and photochemic characterization of a deep ocean proteorhodopsin. *J Biol Chem* **278:** 33985-33991.

Wollenweider RA (1974). Primary Production in Aquatic Environments. Blackwell Scientific Publications: Oxford.

SUPPORTING INFORMATION

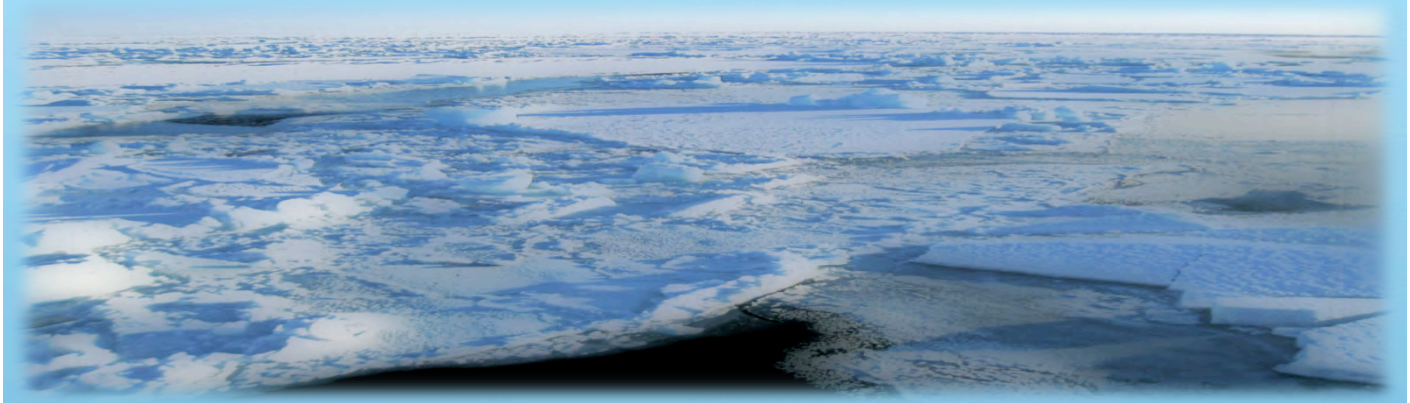
Table SM1. Summary of all the experiments carried out with MED152 and MED134 strains at several concentrations of dissolved organic carbon in the light or in the dark. There are also represented some of the parameters measured: growth rate (μ), doubling time (D_t), lag, and yield.

Exp.	Str.	Medium	DOC [mM]	L/D	μ (h^{-1})	D_t (h)	Lag (h)	Yield cells	CO_2 fixation
Glc+YE	134	Defined	0.03	L	0.002	332.07	0	3.71E+04	-
Glc+YE	134	Defined	0.03	D	0.001	149.36	0	1.80E+04	-
Glc+YE	134	Defined	0.04	L	0.042	11.84	0	7.74E+04	-
Glc+YE	134	Defined	0.04	D	0.027	12.22	0	1.43E+05	-
Glc+YE	134	Defined	0.05	L	0.050	10.62	0	2.08E+05	-
Glc+YE	134	Defined	0.05	D	0.047	10.44	0	2.65E+05	-
Glc+YE	134	Defined	0.06	L	0.051	9.69	0	4.54E+05	-
Glc+YE	134	Defined	0.06	D	0.061	9.71	0	7.18E+05	-
Glc+YE	134	Defined	0.07	L	0.080	8.97	0	6.88E+05	-
Glc+YE	134	Defined	0.07	D	0.064	8.76	0	1.01E+06	-
Glc+YE	134	Defined	0.12	L	0.087	5.26	0	2.36E+06	-
Glc+YE	134	Defined	0.12	D	0.085	5.57	0	2.42E+06	-
Glc+YE	134	Defined	0.73	L	0.140	4.48	0	1.52E+07	-
Glc+YE	134	Defined	0.73	D	0.135	4.32	0	2.30E+07	-
Glc+YE	134	Defined	0.9	L	0.160	3.88	0	4.94E+07	-
Glc+YE	134	Defined	0.9	D	0.153	3.47	0	5.19E+07	-
Glc+YE	134	Defined	250	L	0.145	7.10	0	1.83E+08	-
Glc+YE	134	Defined	250	D	0.143	7.19	0	2.35E+08	-
Glc+YE	152	Defined	0.03	L	0.002	344.33	0	7.33E+04	-
Glc+YE	152	Defined	0.03	D	0.005	665.89	0	4.56E+04	-
Glc+YE	152	Defined	0.04	L	0.059	16.34	0	1.01E+06	-
Glc+YE	152	Defined	0.04	D	0.057	25.37	0	1.09E+06	-
Glc+YE	152	Defined	0.05	L	0.065	13.93	0	2.12E+06	-
Glc+YE	152	Defined	0.05	D	0.066	14.90	0	2.20E+06	-
Glc+YE	152	Defined	0.06	L	0.072	13.63	0	2.71E+06	-
Glc+YE	152	Defined	0.06	D	0.071	11.35	0	2.83E+06	-
Glc+YE	152	Defined	0.07	L	0.077	8.63	0	3.10E+06	-
Glc+YE	152	Defined	0.07	D	0.079	10.90	0	3.71E+06	-
Glc+YE	152	Defined	0.12	L	0.132	7.96	0	5.24E+06	-
Glc+YE	152	Defined	0.12	D	0.124	8.19	0	5.42E+06	-
Glc+YE	152	Defined	0.73	L	0.155	4.95	0	3.41E+07	-

Table SM1. *cont.*

Glc+YE	152	Defined	0.73	D	0.160	5.15	0	2.37E+07	-
Glc+YE	152	Defined	0.9	L	0.178	4.34	0	4.43E+07	-
Glc+YE	152	Defined	0.9	D	0.200	4.54	0	5.84E+07	-
Glc+YE	152	Defined	250	L	0.098	4.78	0	1.02E+08	-
Glc+YE	152	Defined	250	D	0.096	4.86	0	9.35E+07	-
Pept+YE	152	Zobell	0	L	-	-	-	6.50E+04	-
Pept+YE	152	Zobell	0	D	-	-	-	6.39E+04	-
Pept+YE	152	Zobell	0.11	L	-	-	-	2.52E+06	-
Pept+YE	152	Zobell	0.11	D	-	-	-	1.35E+06	-
Pept+YE	152	Zobell	0.37	L	-	-	-	8.06E+06	-
Pept+YE	152	Zobell	0.37	D	-	-	-	5.18E+06	-
Pept+YE	152	Zobell	0.72	L	-	-	-	1.67E+07	-
Pept+YE	152	Zobell	0.72	D	-	-	-	1.08E+07	-
Pept+YE	152	Zobell	0.91	L	-	-	-	2.22E+07	-
Pept+YE	152	Zobell	0.91	D	-	-	-	1.92E+07	-
Pept+YE	152	Zobell	242	L	-	-	-	3.00E+08	-
Pept+YE	152	Zobell	242	D	-	-	-	3.04E+08	-
CO_2 fixation	134	Defined	24.5	L	-	-	-	-	+++
CO_2 fixation	134	Defined	24.5	D	-	-	-	-	++
CO_2 fixation	152	Defined	24.5	L	-	-	-	-	+++
CO_2 fixation	152	Defined	24.5	D	-	-	-	-	++
CO_2 fixation	134	Defined	0.1	L	-	-	-	-	+
CO_2 fixation	134	Defined	0.1	D	-	-	-	-	+
CO_2 fixation	152	Defined	0.1	L	-	-	-	-	++++
CO_2 fixation	152	Defined	0.1	D	-	-	-	-	++

Glc: glucose; **YE:** yeast extract, **Pept:** peptone



Chapter 5

Seasonal changes in substrate utilization patterns by bacterioplankton in the Amundsen Gulf (Western Arctic)

Beatriz Fernández-Gómez, M. Montserrat Sala, and Carlos Pedrós-Alió

ABSTRACT

During the Circumpolar Flaw Lead (CFL) System Study, 17 stations were sampled in the Amundsen Gulf, a dynamic area with heterogeneous ice conditions. In order to track metabolic diversity of heterotrophic bacterioplankton, Biolog MT2 MicroPlates® were inoculated with seawater. Samples were taken from February to July 2008 at the surface (0-12 m) and intermediate depths (20-70 m). In winter the number of substrates used was very low. In early spring, when the levels of chlorophyll *a* increased, so did the number of substrates used. This was followed by a one-month period with no substrates used in April and May. Finally, the activity recovered towards the summer. Carbohydrates were the most commonly used substrates, specially glucose and cellobiose. Amino acids were occasionally used both in surface waters and at intermediate depths. These results were compared to those of a previous study in Franklin Bay, a relatively close area but with a different ice regime, showing a stable layer of land-fast ice throughout the period of study. The differences found in the metabolic capabilities between the two studies were likely a consequence of the different hydrography.

INTRODUCTION

The Arctic Ocean is under the immediate threat posed by global warming in Polar regions. Extensive losses of sea-ice have occurred and a reduction of 42% in ice cover has taken place in the last 30 years due to the increase in temperature (Maslanik *et al*., 2007). Warming, melting, and evaporation observed in the High Arctic during summer 2008 points out the vulnerability of polar microbial ecosystems to ongoing climate change (Vincent *et al*., 2009).

Prokaryotic heterotrophs play a key role in global carbon fluxes through their consumption of dissolved organic matter and nutrient recycling activities (Azam *et al*., 1983; Elifantz H. *et al*., 2007; Pomeroy, 1974). Reduction of the ice cover will affect planktonic communities, particularly modifying timing and intensity of phytoplankton blooms (Carmack *et al*., 2004). These changes could alter the food web structure and the relative importance of top-down versus bottom-up control of marine ecosystems (Arrigo and van Dijken, 2004). Since bacterial production is correlated with primary production (Cole *et al*., 1988), if changes to the phytoplankton occur, they would cascade to secondary production. At small scales, recent evidence suggests that more sea-ice dissolved organic matter (DOC) and exopolysaccharides (EPS) are being released per unit of algal carbon under low rather than high snow cover (Riedel *et al*., 2008). This relationship underlines the importance of studying the Arctic heterotrophic bacterioplankton, since it is the main consumer of dissolved organic matter (Allen, 2002; Cota, 1990).

The winter is the least known part of the year in the Arctic Ocean. The first seasonal study of pelagic Arctic bacterioplankton under the ice was carried out over the Central Arctic Ocean (Sherr and Sherr, 2003; Sherr *et al*., 2003). These authors measured bacterial abundance and heterotrophic production. However, bacterial metabolic diversity under the Arctic ice was not studied. A second seasonal dynamics study determining biomass stocks and productivity of prokaryotes, was carried during the Canadian Arctic Shelf Exchange Study (CASES) in the ice-covered Franklin Bay, on board the CCGS *Amundsen* (Garneau *et al*., 2008). During this cruise, metabolic capabilities of bacterioplankton were studied along an annual cycle (Sala *et al*., 2008). Using Biolog EcoPlates® the latter authors found that polymers and carbohydrates were actively consumed during winter, when chlorophyll *a* concentration was very low, but a lot less in the spring, after the phytoplankton bloom.

In 2008, the International Polar Year (IPY) provided a great opportunity to study the functional characteristics of the microbial communities in the Amundsen Gulf through

the Circumpolar Flaw Lead (CFL) System Study, again on board the CCGS *Amundsen*. The icebreaker navigated through the Cape Bathurst Polynya, an area of open water surrounded by sea ice. The Cape Bathurst Polynya is formed on the Mackenzie Shelf as a result of high winds advecting annual ice away form the land fast ice (Carmack and Macdonald, 2002). Consequently, this area exhibits a marked inter-annual variability in sea ice dynamics that may cause changes in the phytoplankton blooms, both in terms of relative intensity and timing (Arrigo and van Dijken, 2004). Thus, this area provided a more dynamic ecosystem than the stable fast-ice of Franklin Bay studied previously.

Garland and Mills (1991) first applied the commercially available Biolog microplate technique (Biolog) for ecological purposes. This provided a rapid way to assay carbon source utilization by microbial assemblages. Later, Insam (1997) introduced the Biolog Ecoplates, containing carbon sources that were more relevant for natural environments than the existing plates, to characterize microbial assemblages according to their use of distinct carbon sources. Unlike Biolog Ecoplates, Biolog MT2 microplates do not have carbon sources and contain only tetrazolium violet as a redox indicator to reveal the utilization of the carbon sources (Bochner, 1989). Biolog MT2 microplates allow the user to test selected carbon sources according to particular needs (Haack et al., 1995; Preston-Mafham *et al*., 2002). Although MT2 microplates have been used mainly to analyze microbial communities from soil (Campbell, 1997; Zak *et al*., 1994), it has also been employed to asses utilization of carbon sources in freshwater (Fulthorpe and Allen, 1994; Manage *et al*., 2009; Yu and Yu, 2000) and only once in marine environments (Rodríguez *et al*., 2007).

In the present study, we investigated the response of Arctic bacterioplankton to the addition of selected single carbon sources. For this purpose, we chose 23 substrates that were added in triplicate wells. The set of carbon sources included 12 also present in Biolog Ecoplates and 11 new ones. These new carbon sources were: sucrose, maltose, pullulan (polysaccharide polymer consisting of maltotriose units), urea, dimethylsufoniopropionate (DMSP) and its derivatives dimethyl sulfoxide (DMSO) and sodium acrylate, betaine monohydrate, and amino acids such as L-valine, L-leucine, L-lysine, and aspartate.

Our objective was to explore the potential utilization of these carbon sources by bacterioplankton at different depths and periods of the year, and compare the results with those from a previous study in a more stable fast-ice area (Sala *et al*., 2008).

MATERIAL AND METHODS

Cruise plan and sampling

The study was carried out within the framework of the Circumpolar Flaw Lead System Study (CFL: www.ipy-cfl.ca), as part of the International Polar Year and aboard the Canadian icebreaker CCGS *Amundsen*. Samples were taken from February to July 2008 at 17 stations in the Amundsen Gulf plus one in the northeast of Banks Island (74°19'N, 127°1'W) (Fig. 1 and Table 1).

Table 1. Stations occupied with geographical location, date and depths sampled.

Station	Lat. (N)	Long. (W)	Date	Depths (m)		
				Surface	I.D.[a]	Botton
2008-19D	71° 05'	124° 49'	Feb 03	-[b]	30	160
2008-19D	71° 05'	124° 49'	Feb 05	ice[c]	-	-
2008-19D	71° 04'	124° 47'	Feb 11	-	60	160
2008-22D	71° 18'	124° 29'	Feb 18	-	50	-
2008-22D	71° 18'	124° 29'	Feb 19	ice	-	-
2008-26D	70° 56'	123° 55'	Feb 25	-	60	200
2008-26D	70° 56'	123° 55'	Feb 26	ice	-	-
2008-27D	70° 47'	123° 04'	Mar 02	-	70	200
2008-27D	70° 47'	123° 04'	Mar 03	open water[d]		
2008-29D	71° 02'	123° 55'	Mar 10	-	65	200
2008-29D	71° 02'	123° 55'	Mar 11	ice	-	-
2008-33D	71° 04'	121° 47'	Mar 26	12	30, 60	-
2008-36D	71° 11'	124° 05'	Apr 06	10	60	200
2008-38D	71° 14'	124° 37'	Apr 12	10	25	225
2008-41D	70° 47'	122° 20'	Apr 16	ice	30, 75	-
2008-43D	70° 34'	127° 35'	Apr 29	ice	-	-
2008-43D	70° 34'	127° 35'	Apr 30	-	10	153.5
2008-405b	70° 39'	122° 53'	May 19	12	30, 50	-
2008-9008	74° 19'	127° 01'	May 27	12	40, 55	-
2008-F6	69° 51'	123° 45'	Jun 02	10	20, 40	-
2008-FB1	69° 59'	125° 51'	Jun 14	ice	22, 30	-
2008-FB3	69° 58'	125° 53'	Jun 14	ice	20, 30	-
2008-FB5	69° 57'	125°53'	Jun 15	12	17, 30	-
2008-416	71° 28'	127° 75'	Jul 10	ice	57.9, 70	-

[a]Intermediate depths.
[b]Depth not sampled.
[c]ice refers to water collected from just underneath the ice, from a depth of betwee and 2 meters.
[d]open waters refers to water collected in an area without ice cover, from a de between 0 and 2 meters.

During the winter, a layer of ice covers this region. At variable intervals the pack ice breaks away from land-fast ice, creating a flaw lead directly south of Banks Island, widening and forming the Cape Bathurst polynya (71°N, 127°W), a region with very active sea ice dynamics (Arrigo and van Dijken, 2004). Two different depths were analyzed in each sampling: surface (1-12 m) and intermediate depths (20-70 m). At some stations, bottom waters (200-230 m) were also collected. During the winter, samples deeper than 10 m were collected with a CTD connected to a rosette through the moon pool of the ship. In order to avoid ship contamination, surface water (approximately 1-2 m below the ice-water interface) was collected with Niskin bottles through a hole drilled in the ice a few hundred meters away from the ship, assuming that the water mass was the same but without the influence of the ship. When the ice melted, the rosette was deployed overboard.

Chlorophyll *a*

For determination of chlorophyll *a* concentrations 500 ml of seawater was filtered onto 25 mm diameter Whatman GF/F filters and subsequently extracted in 90% acetone in the dark, at 4°C, for 24h, and later determined fluorometrically (Yentsch and Menzel, 1963).

Carbon source utilization

Carbon source utilization was determined using Biolog MT2 plates. These are microtiter plates in which the wells only contain a buffered nutrient medium and the redox indicator dye tetrazolium violet (Biolog Inc.). The Biolog MT plates designed specifically for our study contained 23 different carbon sources in triplicates and three wells received only water as controls (Table 2). Five µl substrate aliquots were added to each of the wells. The carbon compound concentration was set to be 10 mM C final concentration. After sampling, 145 µl of water sample were added to each well and incubated in the dark, at 4ºC, until development of color, usually four weeks. Results were read every week during one month and afterwards they were frozen at –20°C until return to the laboratory, where readings were checked again. A substrate was considered to be positive when at least one of the three wells showed color development. In 133 out of 201 positive results, two or three wells from the triplicates displayed color. Differences between water layers or periods were tested using a t-test with the software Statistica.

Table 2. List of carbon sources added to the Biolog MT2 microplates.

Compounds	Number of positive wells
Polymers	
Glycogen	11
Pullulan	0
Tween 40	1
Tween 80	1
Carbohydrates	
Maltose	12
Sucrose	13
D-Cellobiose	13
Glucose	8
N-acetyl-D-glucosamine	9
Amino acids	
L-Arginine	4
L-Leucine	4
L-Phenylalanine	3
L-Valine	3
L-Threonine	5
L-Lysine	3
Aspartate	1
Other substrates	
DMSP	0
DMSO	1
Acrylate	0
Betaine	0
Urea	0
Pyruvate	3
ATP	0

RESULTS

A total of 35 samples were collected in seventeen stations in the Amundsen Gulf and the eastern Beaufort Sea (Fig. 1). Out of the 23 substrates tested, 17 were used at least on one occasion. The substrates that never showed a positive result were pullulan, DMSP, acrylate, betaine, urea, and ATP. In general, carbohydrates and amino acids were the categories with the highest number of positive results (Fig. 2). The most often used substrates were glucose, maltose, sucrose, glycogen, D-cellobiose and N-acetyl-D-glucosamine. Although the differences were not statistically significant, surface bacteria utilized more substrates than bacteria at deeper depths. Similarly, most of the substrates had a higher number of positive

results at the surface, but this was only significant for glucose (t-test; P=0.022) and for the whole set of amino acids taken together (P=0.036), but not for any amino acid individually. Four substrates (aspartate, Tween 40, Tween 80 and pyruvate) showed positive results only at the surface, whereas DMSO showed positive results only at intermediate depths.

Chlorophyll *a* concentrations at the surface varied between 0.02 and 0.35 µg l^{-1} from February to the beginning of April. As the ice started thinning and melting, chlorophyll *a* increased rapidly, reaching values between 0.40 and 0.71 µg l^{-1} towards the end of the study (Fig. 3). Maximal values were 5.3 and 10.6 µg l^{-1} during the summer. Bacterial abundance did not show clear differences over winter and early spring, showing constant values between approximately 1 and $5x10^5$ bacteria ml^{-1}. However, bacteria started to increase from July on, reaching a maximum of $1.4x10^6$ bacteria ml^{-1} (Nguyen and Maranger, submitted).

The number of substrates used over the entire period showed a similar trend to that of chlorophyll *a* concentrations. In February, when the concentrations of chlorophyll *a* were

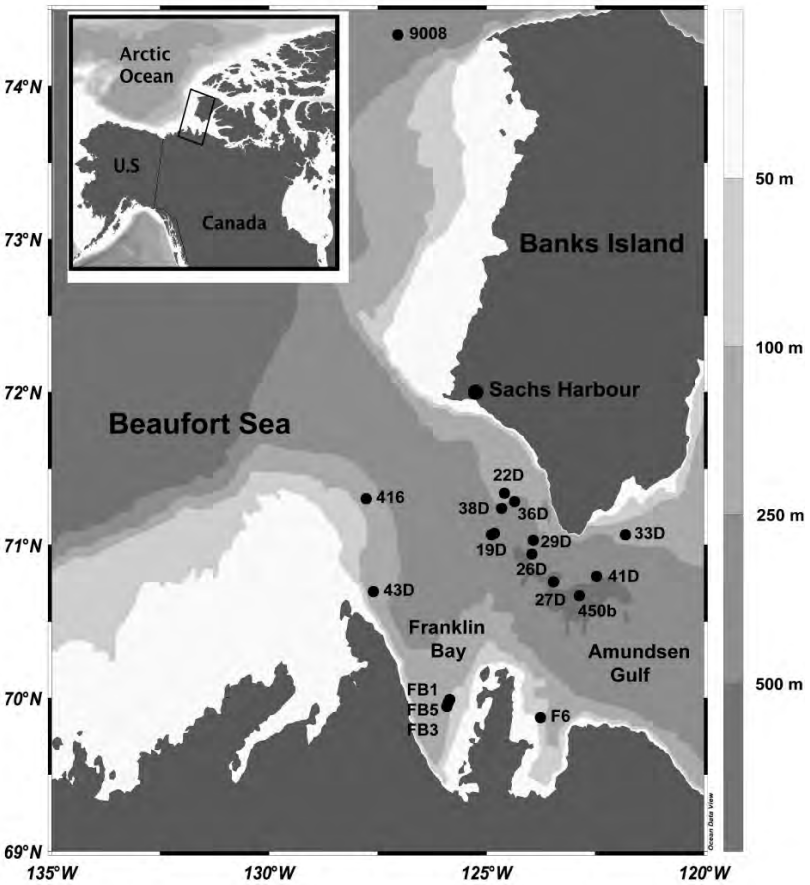


Figure 1. Stations sampled in the Amundsen Gulf and the Beaufort Sea. Stations FB1, FB3 and FB5 were in Franklin Bay. Shadows of gray indicate bathymetry (see vertical scale bar).

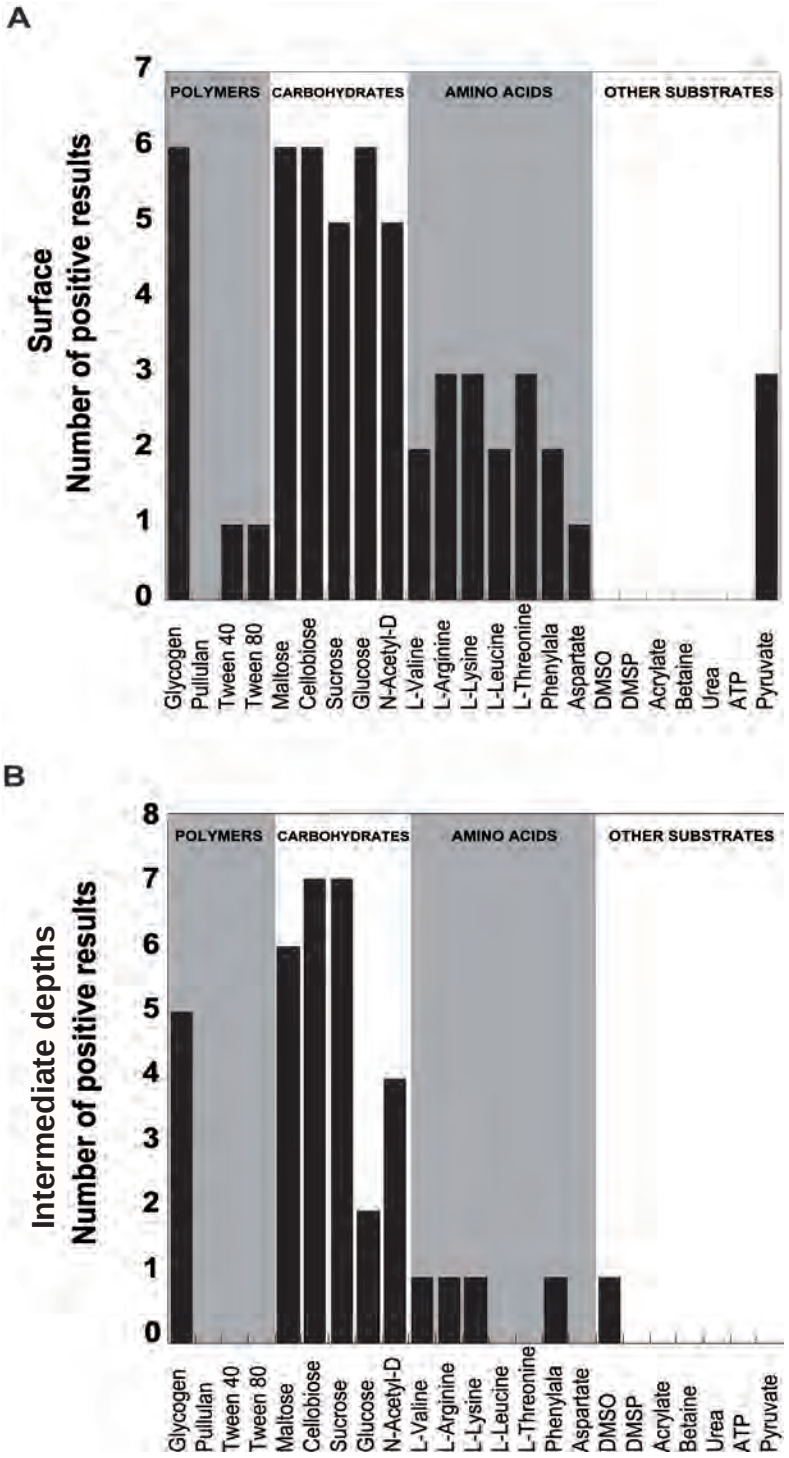


Figure 2. Number of positive results for substrates in the Biolog MT2 microplates. Substrates have been grouped into categories and results are shown for the two different layers: **(A)** surface and **(B)** intermediate depths.

low, utilization of the substrates could not be detected. Some substrates started to be used in March, after chlorophyll *a* started to increase, with the exception of the period from the end of April to the beginning of May, when there were no positive results (Fig. 3). Utilization of all categories of substrates was low in winter. Only L-leucine showed significantly more positive results in winter than in spring (P=0.004).

At intermediate depths chlorophyll *a* concentrations were very low, with the highest value at the end of May (3.3 µg l^{-1}, Fig. 3). Very few substrates were used in winter and none during the peaks of chlorophyll *a* at the end of April and the beginning of May. This lack of positive

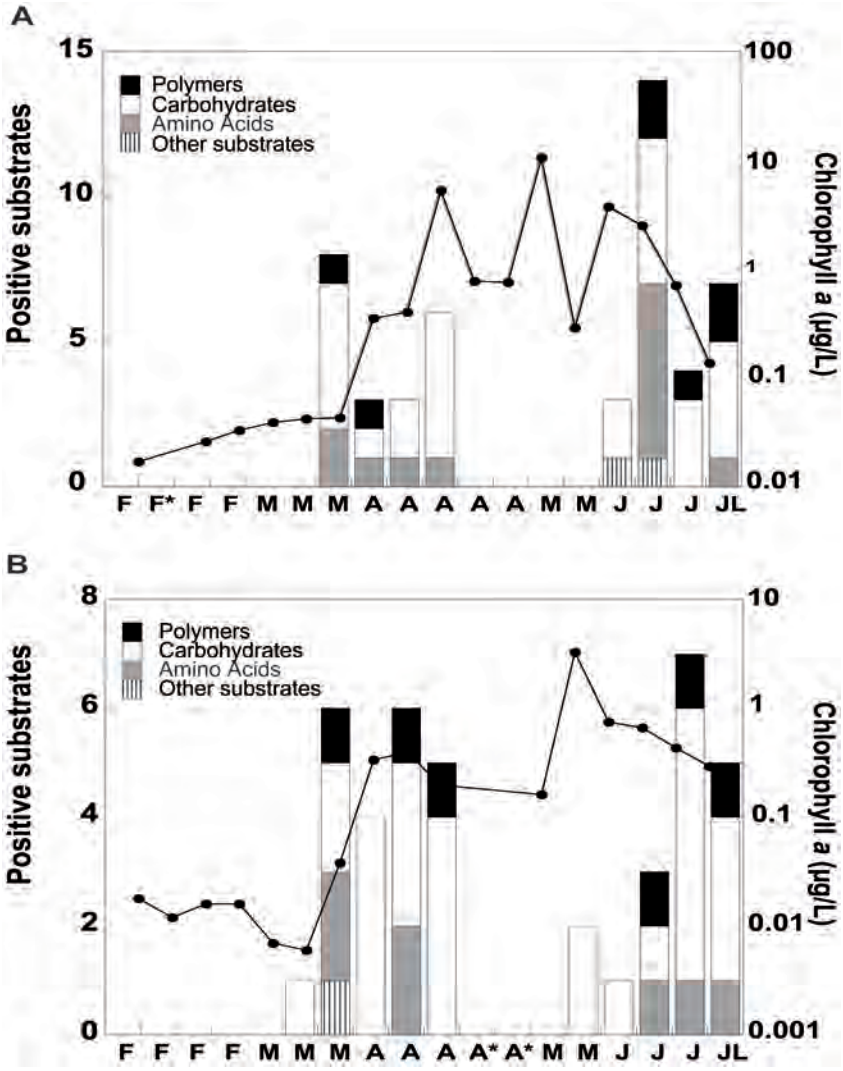


Figure 3. Chlorophyll *a* concentration (line and filled circles), and number of substrates used (bars) separated into substrate categories, for the surface samples **(A)** and intermediate depths **(B)** along the period of study.
The asterisk (*) at some stations indicates that no sample was collected at that particular spot. Note different vertical scales at the two depths.

responses in late April and early May was, thus, observed both at intermediate depths and at the surface. The largest number of substrates used was found in March-April and June-July. Following the same pattern as at the surface, carbohydrates were the main substrates used, followed by amino acids (Fig. 2). Bacterial counts in this layer followed a similar pattern to that at the surface although numbers were lower. Values were very low during February-May ranging between $6x10^4$ and $4x10^5$ bacteria ml^{-1}, and increased from July on, reaching up to $8x10^5$ bacteria ml^{-1} (Nguyen and Maranger, submitted).

The eight substrates with largest number of positive results were chosen for a detailed analysis (Fig. 4). At the surface, most of the substrates were used quite regularly from the end of March until the end of the study, except for one month (end of April beginning of May) when there was no utilization at all. L-leucine was the only compound that showed positive results in late spring only, both at the surface and the intermediate layer.
Taking both depths together, most of the substrates were used more in spring than in winter. However, a t-test revealed no significant differences between winter and spring, except for L-leucine (P=0.004).

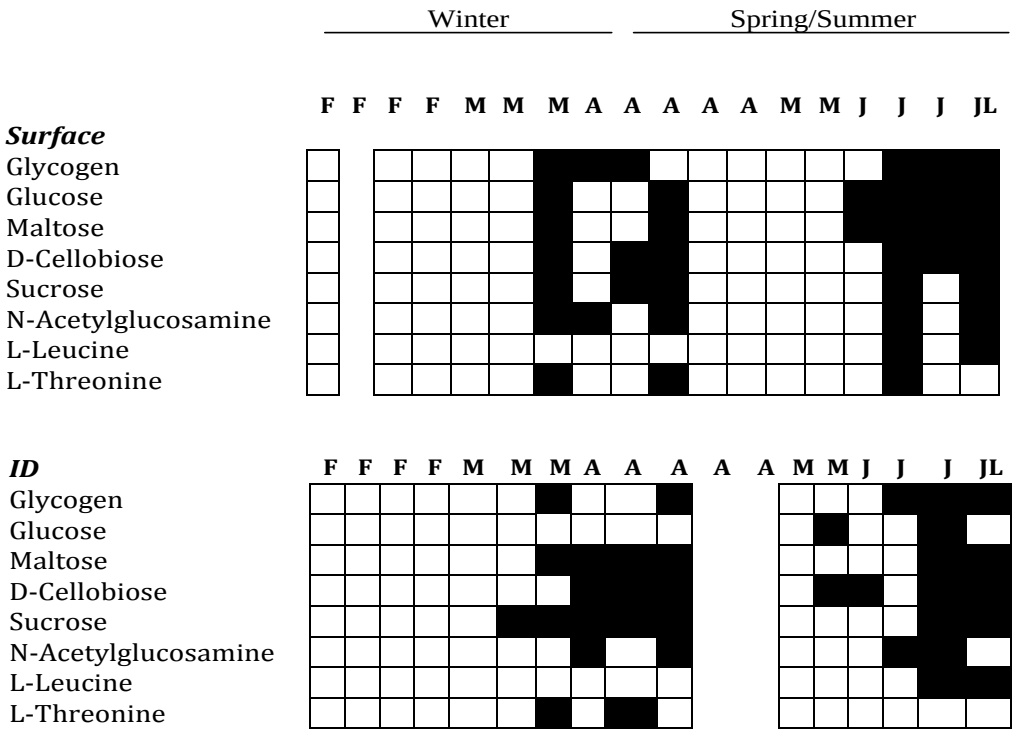


Figure 4. Pattern of utilization of the eight most used substrates in the Biolog MT2 microplates along the study. Letters denote month of the year. A positive utilization of the substrate is marked as a filled square whereas no utilization is marked as an empty square. Missing columns indicate that there was no sample for that date.

We compared the eight compounds that were present in both CASES and CFL studies (Fig. 5A). Generally the percent of times used was similar for most of the substrates, but the largest differences were found for Tween 40 and Tween 80, that were used more times in CASES than CFL. On the other hand, amino acids were used in CFL more often than in CASES. Glycogen, N-acetyl-D-glucosamine and D-cellobiose were frequently used both in CFL and in CASES. The few samples that were collected from deeper depths (200-230 m) produced no positive results (data not show) and they were not considered in the analysis.

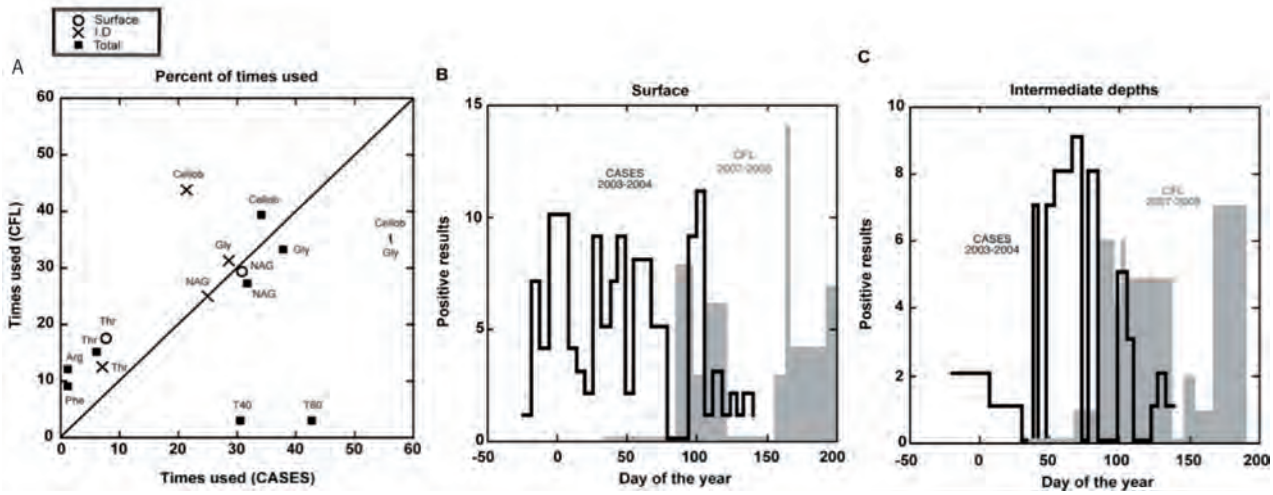


Figure 5. A) Percentage of utilization of various compounds studied in the CASES cruise (2003-2004) compared with the present study. Each compound is indicated by three points: surface, intermediate depths, and total. The diagonal shows the 1 to 1 line.
NAG: N-acetyl-D-glucosamine; **Gly**: glycogen; **Cellob**: cellobiose; **T40**: Tween 40; **T80**: Tween 80; **Thr**: threonine; **Phe**: phenylalanine; **Arg**: arginine.
B) Number of positive results along time in the two studies CASES (solid line) and present study (grey area) at the surface. Note horizontal lines indicating zero positive results for the initial dates of the CFL study. **C)** Same as **B)** for intermediate depths.

DISCUSSION

Sala *et al*. (2008) conducted a study of the metabolic diversity of bacterioplankton using Biolog Ecoplates in Franklin Bay from winter 2003 to spring 2004. These authors found that carbohydrates, especially polymers, were the compounds most commonly used throughout the period studied. Thus, we decided to pay special attention to these compounds. We explored the possibility of using several different carbohydrates in Biolog MT2 plates. In principle, we wanted to use as many different possible polymers of natural origin as possible. Particularly, we were interested in having both storage and structural carbohydrates from

both phototrophic (plants and algae) and heterotrophic (bacteria and animals) origins. The main compounds of interest are summarized in Table 3.

Table 3. Carbohydrate family compounds targeted and substrates used in the study by Sala *et al.* (2008) using Biolog Ecoplates and in the present study, using Biolog MT2 plates.

Target compound	Substrate tested	Number of positive results (%)	
		Sala *et al*. (2008)	Present study
Phototrophic origin			
Cellulose	Cellobiose	28 (37.3)	13 (39.3)
Starch	Pullulan	nt[a]	0
	Maltose	nt	12 (36)
	α-cyclodextrin[b]	30 (40)	nt
Hemicellulose	D-Xylose	10 (13.3)	nt
Heterotrophic origin			
Chitin and Peptidoglycan	NAG[c]	26 (34.7)	9 (27.2)
Glycogen	Glycogen	31 (41.3)	11 (33)
General			
Glucose-1P	Glucose-1P	1 (1.3)	nt
Glucose	Glucose	nt	8 (24)
Sucrose	Sucrose	nt	13 (39.3)
Total assays		**75**	**33**

[a]nt: not tested; [b] starch analog; [c]NAG: N-acetyl-D-glucosamine.

Initially, we targeted cellulose and starch from phototrophic origin and chitin and glycogen from heterotrophic origin. However, some of these polymers precipitated when added at the necessary concentration (10 mM). Such high concentrations are necessary to obtain enough bacterial growth in the Biolog plate for the indicator to change color. Of all the polymers assayed, cellulose, starch, xylan, agar, gelatin and chitin proved to be unsuitable for this reason. Only glycogen and pullulan remained in solution at this concentration. For the other compounds we had to resort to oligomers resulting from their degradation (Table 3). In effect, maltose results from starch degradation, cellobiose from cellulose, and N-acetyl-D-glucosamine from both chitin and bacterial peptidoglycan. All these compounds were appropriate for our purpose. In addition, we also tested glucose and sucrose, as two widely distributed, simple, and easily assimilated sugars. Except for pullulan (see below) these compounds were used the most times in agreement with the study of Sala *et al*. (2008). According to the number of positive results found, it did not appear to be any preference for one of these compounds over the others. All of them seemed to be equally used for growth. In

fact, the number of positive results was essentially the same as those obtained with sucrose and glucose. Thus, the Arctic assemblage seemed well prepared to degrade most polymeric carbohydrates irrespectively of their origin.

Pullulan was the only carbohydrate polymer that did not show any positive result. This polysaccharide was selected because it originates from marine sources and pullulanase has been detected in marine bacteria (Arnosti *et al*., 2005; Arnosti and Repeta, 1994). Arnosti *et al*. (2005) and Arnosti (2008) found that among all the extracellular enzyme activities they tested in the Svalbard archipelago, pullulan was hydrolyzed in sediments but not in seawater. Our results agree with this, but the possible reason remains unknown.

Amino acids were the second group of compounds in number of positive results. Although the total number of positives was moderate, all of the amino acids tested provided some positive results. We tried to use gelatin as a protein polymer but, again, it precipitated at the necessary concentrations. The evidence for amino acid utilization in cold waters is somewhat contradictory. Thus, Pomeroy *et al*. (1990) found low amino acid uptake by prokaryotic assemblages from Arctic water in spring. Sala *et al*. (2008), in the CASES study, also found a very low number of positive results in the wells with amino acids. Glycyl-L-glutamic acid was the only amino acidic compound that was used significantly. Sala et al. (2005) detected very few positive results with amino acids in western Antarctic waters in summer. On the other hand, Jellet *et al*. (1996) found intense use of amino acids in the Bedford Basin. Likewise, studies carried out with microautoradiography combined with fluorescence in situ hybridization (MARFISH) in Arctic waters showed the highest percentage of cells labeled with amino acids both in summer (Elifantz *et al*., 2007) and winter (Alonso-Saéz *et al*., 2008). This is intriguing because the studies of Sala *et al*. (2008) and Alonso-Sáez *et al*. (2008) were carried out simultaneously on the same cruise. Apparently, many bacteria were able to take up the amino acids, but they were not able to grow enough to change the color of the indicator in Biolog plates. In the present study with Biolog plates, amino acid use was significant, suggesting that, unlike in the case with carbohydrates, amino acid use in the Arctic is quite variable in different areas and seasons.

The third group of compounds tested was a heterogeneous mixture of compounds related to microbial compatible solutes. Ice algae must be an important source of organic matter during the winter and spring. Since this community is subject to osmotic stress, and compatible solutes are produced in large quantities, it seemed reasonable to test whether the bacterial assemblage was adapted to degrade some of these compounds or not. DMSP is produced by algae (Karsten *et al*., 1990; Keller *et al*., 1989) and acts as a cryoprotectant and

osmoregulator (Kirst *et al*., 1991; Lee *et al*., 2001; Tan, 1997). It is abundant in polar waters (Lee *et al*., 2001) and it is an important substrate for a broad and diverse fraction of marine bacterioplankton (Vila-Costa *et al*., 2004). In the CASES cruise, Vila-Costa *et al*. (2008) found that around 5% of the bacteria were active in assimilating DMSP using MARFISH. Thus, we tested DMSP as well as two compounds that result from its processing by microbial communities: acrylate and DMSO. Finally, we also tested another compatible solute: betaine. None of these compounds produced positive results. We hypothesize that even though DMSP is likely an important substrate for the bacterioplankton, neither DMSP nor its derivatives are enough as sole carbon sources for growth on Biolog plates.

Two of the products tested were Tween 40 and Tween 80. These compounds are present in Biolog Ecoplates and have been useful in the separation of natural microbial communities (Hitzl *et al*., 1997). Presumably, the lipid chain is a test for respiration of lipids and the polar part of the molecule allows it to be used in aqueous media. Both compounds were among the most used substrates in previous studies carried out in Polar areas of western Antarctica (Sala *et al*., 2005), the Canadian Arctic (Sala *et al*., 2008), and the Greenland Sea (Frette *et al*., 2010; Sala *et al*., 2010; Tan, 1997). In the present study, however, Tween 40 and 80 only showed a couple of positive results and this was one of the most striking differences with the CASES study (Sala *et al*., 2008).

We included ATP because Alonso-Sáez *et al*. (2008) found a substantial fraction of the bacteria labeled with this compound with the MARFISH technique. However, we found no positive results during the present study. The same was true for urea. This compound was included because previous studies in the area (Estrada *et al*., 2009) had shown that organic nitrogen was likely important for the microbial assemblage. Urea accounts for a large percentage of the total dissolved nitrogen in the eastern Canadian Arctic (Harrison *et al*., 1985). Obviously, these compounds did not elicit enough respiration to change the color of the indicator in the Biolog plates.

The study of Sala *et al*. (2008) had many similarities to the present one. It was carried out in Franklin Bay, while the present study was carried out a little to the north in the Amundsen Gulf (and we included a few stations in Franklin Bay) (data not shown). Both studies included a several month period from winter to spring and the Biolog plates used had several substrates in common. In fact, the present study was undertaken to test some of the hypotheses derived from the previous study. The main difference among the studies was the ice dynamics of the region sampled. Franklin Bay was covered by a very stable layer of landfast ice throughout the period studied. The Amundsen Gulf, on the other hand, was a very

heterogeneous mixture of first year ice and open waters. The ship would be anchored to an ice floe for a few days, drifting with the floe, generally towards the west. Then it would break free of the ice and reposition itself on another floe further to the east. As a result, different stations had very different conditions of ice cover, and the whole region was very dynamic, experiencing open water conditions and mixing at some point or another. This was reflected in the mixed layer depth (MLD). In CASES the MLD was uniformly shallow along the study, while in CFL it was much more variable with several periods of deeper mixing (Barette *et al*., 2010; Nahavandian *et al*., 2010). It seems reasonable to attribute differences in metabolic fingerprints to these different conditions. On the other hand, those patterns that repeated themselves in the two cruises could be taken to be robust features of the Arctic community.

As already discussed, carbohydrates and, particularly polymeric carbohydrates, produced the largest number of positive results in both studies (Fig. 5A). The overall use was similar in both cruises for cellobiose, glycogen and N-acetylglucosamine. The only difference was that these compounds were preferentially used at the surface in CASES and at intermediate depths in CFL. On the contrary, amino acids showed a much larger proportion of positive results in CFL than in cases. And Tween 40 and Tween 80 showed a much larger number of positive results in CASES. These results suggest that carbohydrates are an important substrate in the Arctic winter and spring irrespective of the ice conditions in the area. For the other compounds more studies are needed.

The CASES study started in December and ended in May, while the present study lasted from February to July. Thus, both studies overlapped for a period of four months between February and May. The number of positive results showed marked variations along these time periods (Fig. 5B and C). In the present study we found two periods of high activity, one in April and one in late June and July. The pattern was the same at both depths. These periods of activity coincided with the ice melt in April and the annual phytoplankton bloom in summer. Apparently, these two periods create fast changing conditions that result in more positive results of the bacterioplankton. At both depths, we had two periods where no positive results were found, one in winter and another one in late April and May. Sala *et al*. (2008) also found a period without positive results after the ice melt at all depths studied. These authors reasoned that the change from winter to spring and summer conditions caused a change in the bacterioplankton community composition from a winter-adapted to a summer-adapted assemblage (Fig. 1 SM). The lack of positive results was likely a result of one assemblage disappearing and the new one appearing. The present study shows that this phenomenon occurred in the two different cruises. This indicates that these periods are a robust feature of the Arctic environment.

There is one difference between the two studies remaining. At intermediate depths (Fig. 5C) the patterns of positive results were similar in the two studies, although the periods of activity were slightly displaced in time. This was also the case for the spring to summer patterns at the surface (Fig. 5B). However, the winter pattern at the surface was completely different with very high number of positive results in CASES and zero in the present study (Fig. 5B). We must attribute this difference to the ice conditions in the two areas of study. Presumably, the stable land-fast ice and shallower and less variable mixed layer depth in CASES allowed the accumulation of organic matter originating in the ice community in the water layer immediately below the ice (our surface samples). The changing and frequently mixed conditions in the present study, on the other hand, likely prevented this accumulation of organic matter next to the surface and, therefore, the results were the same as those at intermediate depths.

CONCLUSIONS

The comparison between two different winter-to-summer periods in the Western Arctic suggests that carbohydrates, particularly polymeric molecules, are an important carbon source for the bacterioplankton. Amino acids were also consumed but, as observed in other studies, its use is rather variable depending on the areas and the season. The existence of periods without positive results turned out to be another robust feature of the Arctic assemblage. One of these periods coincided with ice breakout, a time of rapid change in the environment during ice melt. Dissimilarities with CASES were likely due to differences in ice dynamics in both regions that might affect both the metabolic fingerprint of the community and the pattern of activity at the different depths. In particular, a shallower and more stable mixed layer depth under the land fast ice in CASES favored bacterial activities, while the variable conditions experienced during CFL resulted in low activity levels.

ACKNOWLEDGMENTS

We thank Roxane Maranger (Université de Montréal) and Connie Lovejoy (Université Laval) and their teams for sharing lab equipment and help during sampling and experiments on board. We also thank to our on board sampling team formed by Cèlia Marrasé, Raquel Rodríguez-Martínez, Montserrat Coll-Lladó, Laura Alonso-Sáez, Marta Estrada, and Cristina Romera-Castillo. Special thanks to Dan Nguyen who helped during sampling and performed the DGGE analysis. Finally we are very grateful to our fellow scientists on board and to the officers and crew of the CCCG *Amundsen* for their support during the cruise.

This work is a contribution to the International Polar Year-Circumpolar Flaw Lead system study (IPY-CFL 2008) supported through grants from the Canadian IPY Federal Program Office and the National Sciences and Engineering Research Council. The Spanish participation was funded by the Spanish Ministry of Science and Innovation grant BOREAL (CGL2007-28872-E/ANT) to C. P.-A.

REFERENCES

Allen, A., 2002. Importance of heterotrophic bacterial assimilation of ammonium and nitrate in the Barents Sea during summer. Journal of Marine Systems, 38(1-2): 93-108.

Alonso-Sáez, L., Sánchez, O., Gasol, J., Balagué, V. and Pedrós-Alió, C., 2008. Winter-to-summer changes in the composition and single-cell activity of near-surface Arctic prokaryotes. Environmental Microbiology, 10(9): 2444-2454.

Arnosti, C., 2008. Functional differences between Arctic seawater and sedimentary microbial communities: contrasts in microbial hydrolysis of complex substrates. FEMS Microbiology Ecology, 66(2): 343-351.

Arnosti C., Durkin S., and Jeffrey W. H., 2005. Patterns of extracellular enzyme activities among pelagic marine microbial communities: implications for cycling of dissolved organic carbon. Aquatic Microbial Ecology, 38: 135-145.

Arnosti, C. and Repeta, D.J., 1994. Extracellular enzyme activity in anaerobic bacterial cultures: evidence of pullulanase activity among mesophilic marine bacteria. Applied and Environmental Microbiology, 60(3): 840-846.

Arrigo, K. and van Dijken, G., 2004. Annual cycles of sea ice and phytoplankton in Cape Bathurst polynya, southeastern Beaufort Sea, Canadian Arctic. Geophysical Research Letters, 31(8): L08304.

Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyer-Reil, L. A., Thingstad, F., 1983. The ecological role of water column microbes in the sea. Marine Ecology Progress Series, 10: 257-263.

Barette, J., Gratton, Y., and Prieur, L., 2010. Eddies in the Amundsen Gulf, ArcticNet Scientific meeting. Otawa, Ontario.

Biolog, Instructions for use of Biolog MT plates. Biolog Inc. Hayward, California, USA, 1993.

Bochner, B., 1989. “Breathprints” at the microbial level. ASM News, 55: 536-539.

Campbell, C., 1997. Use of rhizosphere carbon sources in sole carbon source tests to discriminate soil microbial communities. Journal of Microbiological Methods, 30(1): 33-41.

Carmack, E.C. and Macdonald, R.W., 2002. Oceanography of the Canadian Shelf of the Beaufort Sea: A setting for marine life. Arctic, 55: 29-45.

Carmack, E.C., Macdonald, R.W. and Jasper, S., 2004. Phytoplankton productivity on the Canadian Shelf of the Beaufort Sea. Marine Ecology Progress Series, 277: 37-50.

Cole, J., Findlay, S. and Pace, M., 1988. Bacterial production in fresh and saltwater ecosystems: a cross-system overview. Marine Ecology Progress Series, 43: 1-10.

Cota, G., 1990. Bacterioplankton in the marginal ice zone of the Weddell Sea: biomass, production and metabolic activities during austral autumn. Deep Sea Research Part

A. Oceanographic Research Papers, 37(7): 1145-1167.

Elifantz H., Dittel A.I., Cottrell M.T. and Kirchman D.L., 2007. Dissolved organic matter assimilation by heterotrophic bacterial groups in the western Arctic Ocean. Aquatic Microbial Ecology, 50: 39-49.

Estrada, M., Bayer-Giraldi M., Felipe J., Marrasé C., Sala M.M., Vidal, M., 2009. Light and nutrient effects on microbial communities collected during spring and summer in the Beaufort Sea. Aquatic Microbial Ecology, 54(3): 217-231.

Frette, L., Winding, A. and Kroer, N., 2010. Genetic and metabolic diversity of Arctic bacterioplankton during the post-spring phytoplankton bloom in Disko Bay, western Greenland. Aquatic Microbial Ecology, 60(1): 29-41.

Fulthorpe, R.R. and Allen, D.G., 1994. Evaluation of Biolog MT plates for aromatic and chloroaromatic substrate utilization tests. Canadian Journal of Microbiology, 40(12): 1067-1071.

Garland, J. and Mills, A., 1991. Classification and characterization of heterotrophic microbial communities on the basis of patterns of community-level sole-carbon-source utilization. Applied Environmental Microbiology, 57(8): 2351-2359.

Garneau, M.-È., Roy, S., Lovejoy, C., Gratton, Y. and Vincent, W., 2008. Seasonal dynamics of bacterial biomass and production in a coastal Arctic ecosystem: Franklin Bay, western Canadian Arctic. Journal of Geophysical Research, 113(C7).

Haack, S.K., Garchow, H., Klug, M.J. and Forney, L.J., 1995. Analysis of factors affecting the accuracy, reproducibility, and interpretation of microbial community carbon source utilization patterns. Applied Environmental Microbiology, 61(4): 1458-1468.

Harrison, W.G., Head, E.J.H., Conover, R.J., Longhurst, A.R. and Sameoto, D.D., 1985. The distribution and metabolism of urea in the eastern Canadian Arctic. Deep Sea Research Part A. Oceanographic Research Papers, 32(1): 23-42.

Hitzl, W., Rangger, A., Sharma, S. and Insam, H., 1997. Separation power of the 95 substrates of the BIOLOG system determined in various soils. FEMS Microbiology Ecology, 22(3): 167-174.

Insam, H., 1997. Substrate utilization tests in microbial ecology. A preface to the special issue of the Journal of Microbiological Methods. Journal of Microbiological Methods, 30(1-2).

Jellet, J., Li, W., Dickie, P., Boraie, A and Kepkay, P., 1996. Metabolic activity of bacterioplankton communities assessed by flow cytometry and single carbon substrate utilization. Marine Ecology Progress Series, 136: 213-225.

Karsten, U., Wiencke, C. and Kirst, G.O., 1990. The ß-dimethylsulphoniopropionate (DMSP) content of macroalgae from Antarctica and Southern Chile. Botanica Marina, 33(2): 143-146.

Keller, M.D., W.K. Bellows, and R.R.L. Guillard, 1989. Dimethyl sulfide production in marine phytoplankton. In: E.S. Saltzman and W.J. Cooper (eds.), Biogenic Sulfur in the Environment. American Chemical Society Symposium Series No. 393, Washington, DC, pp. 167-182.

Kirst, G.O., Thiel, C., Wolff, H., Nothnagel, J., Wanzek, M. and Ulmke, R., 1991. Dimethylsulfoniopropionate (DMSP) in icealgae and its possible biological role. Marine Chemistry, 35(1-4): 381-388.

Lee, P.A., de Mora, S.J., Gosselin, M., Levasseur M., Bouillon, R.C., Nozais, C., Michel, C., 2001. Particulate dimethylsulfoxide in Arctic sea-ice algal communities: the cryoprotectant hypothesis revisited. Journal of Phycology, 37(4): 488-499.

Manage, P., Edwards, C., Singh, B. and Lawton, L., 2009. Isolation and identification of novel microcystin-degrading bacteria. Applied Environmental Microbiology, 75(21): 6924-6928.

Maslanik, J.A., Fowler, C., Stroeve, J., Drobot, S., Zwally, J., Yi, D., Emery, W., 2007. A younger, thinner Arctic ice cover: Increased potential for rapid, extensive sea-ice loss. Geophysical Research Letters, 34.

Nahavandian, S., Gratton, Y. and Prieur, L., 2010. Evolution of the mixed layer in the Amundsen Gulf and Beaufort Sea. Québec-Océan General Annual Meeting, Lac Beauport, Québec.

Pomeroy, L.R., 1974. The ocean's food web, a changing paradigm. BioScience, 24(9): 499-504.

Pomeroy, L.R., Macko, S.A., Ostrom, P.H. and Dunphy, J., 1990. The microbial food web in Arctic seawater concentration of dissolved free amino acids and bacterial abundance and activity in the Arctic Ocean and in Resolute Passage. Marine Ecology Progress Series, 61: 31-40.

Preston-Mafham, J., Boddy, L. and Randerson, P., 2002. Analysis of microbial community functional diversity using sole-carbon-source utilisation profiles-a critique. FEMS Microbiology Ecology, 42(1): 1-14.

Riedel, A., Michel, C., Gosselin, M. and Leblanc, B., 2008. Winter-spring dynamics in sea-ice carbon cycling in the coastal Arctic Ocean. Journal of Marine Systems, 74(3-4): 918-932.

Rodríguez, N., Massol, A., Imam, S.H. and Zaidi, B.R., 2007. Microbial utilization of toxic chemicals in surface waters of Guayanilla Bay, Puerto Rico: Impact of seasonal variation. Caribbean Journal of Science, 43: 172-180.

Sala, M.M., Arin L., Balagué V., Felipe J., Guadayol O., Vaqué D., 2005. Functional diversity of bacterioplankton assemblages in western Antarctic seawaters during late spring. Marine Ecology Progress Series, 292: 13-21.

Sala, M.M., Arrieta, J.M., Boras, J.A., Duarte, C.M. and Vaqué, D., 2010. The impact of ice melting on bacterioplankton in the Arctic Ocean. Polar Biology, 33(12): 1683-1694.

Sala, M.M., Terrado, R., Lovejoy, C., Unrein, F. and Pedrós-Alió, C., 2008. Metabolic diversity of heterotrophic bacterioplankton over winter and spring in the coastal Arctic Ocean. Environmental Microbiology, 10(4): 942-949.

Sherr, B. and Sherr, E., 2003. Community respiration/production and bacterial activity in the upper water column of the central Arctic Ocean. Deep Sea Research Part I: Oceanographic Research Papers, 50(4): 529-542.

Sherr, E., Sherr, B., Wheeler, P. and Thompson, K., 2003. Temporal and spatial variation in stocks of autotrophic and heterotrophic microbes in the upper water column of the central Arctic Ocean. Deep Sea Research Part I: Oceanographic Research Papers, 50(5): 557-571.

Tan, T.L., 1997. Biolog metabolic fingerprints for clustering marine oligotrophic bacteria from polar regions. In: Insam, H and Rangger, A. (eds.), Microbial communities. Functional versus structural approaches. Springer-Verlag, Berlin, Heildelberg, Germany, 161-170 pp.

Vila-Costa, M., Simó, R., Alonso-Sáez, L. and Pedrós-Alió, C., 2008. Number and phylogenetic affiliation of bacteria assimilating dimethylsulfoniopropionate and leucine in the ice-covered coastal Arctic Ocean. Journal of Marine Systems, 74(3-4): 957-963.

Vila-Costa, M., Simó, R., Kiene R.P., Pinhass,i J., González, J.M., Moran, M.A., Pedrós-Alió, C., 2004. Use of microautoradiography combined with fluorescence in situ hybridization to determine dimethylsulfoniopropionate incorporation by marine bacterioplankton taxa. Applied and Environmental Microbiology: 4648-4657.

Vincent, W., Whyte, L., Lovejoy, C., Greer, C., Laurion, I., Suttle, C., Corbeil, J., Mueller, D., 2009. Arctic microbial ecosystems and impacts of extreme warming during the International Polar Year. Polar Science, 3(3):171-180.

Yu, C.-P. and Yu, Y.-H., 2000. Evaluation of biodegradability by the reduction of tetrazolium violet in Biolog microplates. Biotechnology Letters, 22(11): 909-913.

Yentsch, C. S. and Menzel, D. W., 1963. A method of the determination of phytoplankton chlorophyll and phaeophytin by fluorescence. Deep Sea Research, 10: 221-231.

Zak, J.C., Willig, M.R., Moorhead, D.L. and Wildman, H.G., 1994. Functional diversity of microbial communities: A quantitative approach. Soil Biology and Biochemistry, 26(9): 1101-1108.

SUPPORTING INFORMATION

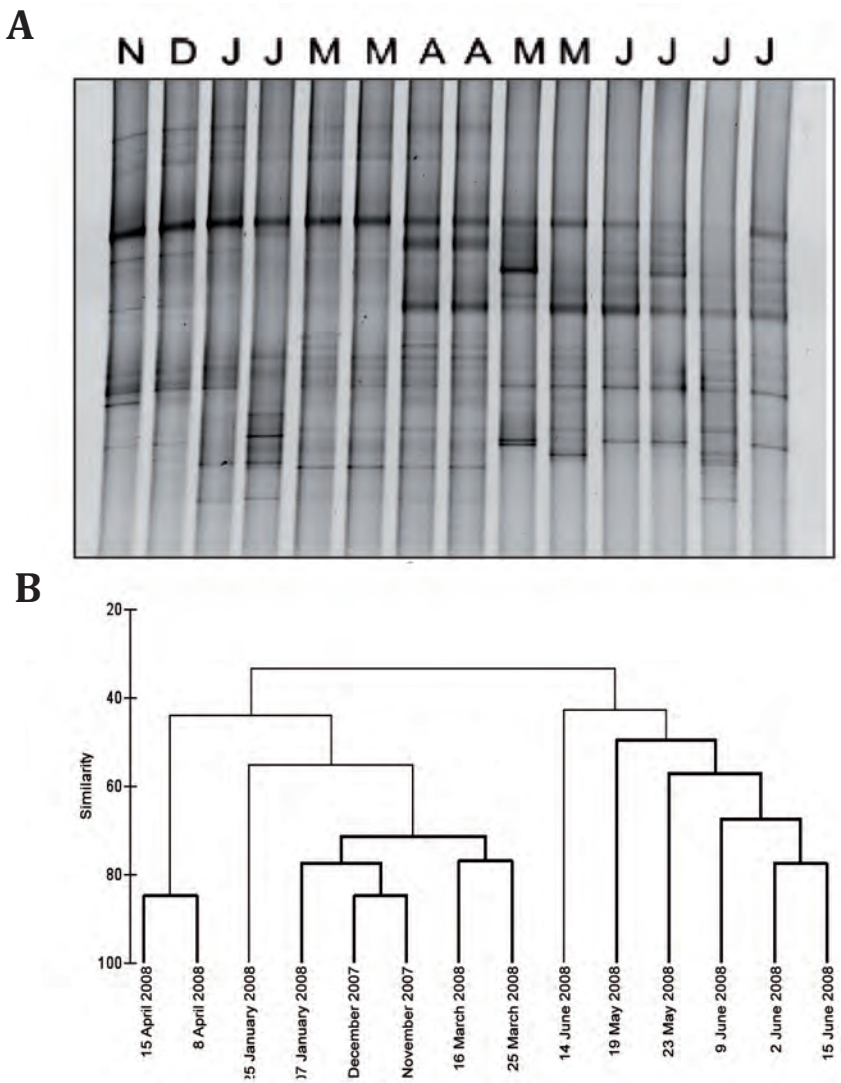


Figure 1SM. (A) 16S rDNA denaturing gradient gel electrophoresis (DGGE) and (B) cluster analysis. DGGE was performed following a protocol modified from Schauer *et al*. (2003). The 16S ribosomal subunit was amplified using bacterial primers GC358F, carrying a GC clamp, and 907RM. PCR products were allowed to migrate at 100V for 17h at 60°C and loaded on a 6% acrylamide gel containing a denaturing gradient of 40 to 80%, where 100% denaturing agent contained 7M urea and 40% deionised formamide. DGGE was carried out using a DGGE 2000 system (CBS scientific). A matrix of the relative contribution of each band to the total intensity of each given lane was built using the Quantity One software (Bio-Rad). From this matrix, a dendrogram based on Bray-Curtis distance was calculated (Dan Nguyen).

Schauer M., Balagué, V., Pedrós-Alió, C. and Massana, R., 2003. Seasonal changes in the taxonomic composition of bacterioplankton in a coastal oligotrophic system. Aquat Microb Ecol, 31(2): 163-174.

Synthesis of Results and General Discussion

SYNTHESIS OF RESULTS AND GENERAL DISCUSSION

Marine Bacteroidetes from a genetic and physiological perspective

The purpose of the present dissertation was to increase the understanding about the genomics, ecology and physiology of the marine Bacteroidetes by using genomic approaches and bacterial cultures. Bacteroidetes are one of the three most abundant groups of bacteria in the ocean (Alonso *et al*., 2007, Alonso-Sáez and Gasol 2007). In recent years, the number of complete bacterial genomes has been exponentially increasing in databases such as GenBank. Currently, the number of Bacteroidetes genomes represents the largest proportion (8.9%) after Proteobacteria (29.4%; gamma- 17.8% and alphaproteobacteria 11.6%) and Firmicutes (22.9%)(Fig. 1A). Besides, the Gordon and Betty Moore Foundation Initiative has sequenced 16 marine Bacteroidetes out of 182 marine bacterial genomes (Fig. 1B.).

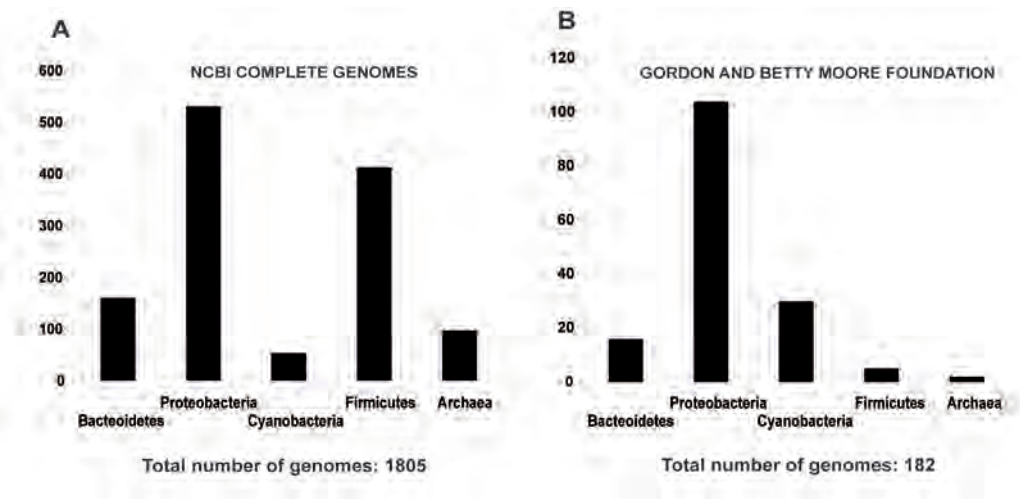


Figure 1. Number of complete genomes of the most abundant groups in NCBI (Genbank) and G&B Moore Foundation (update 07052011).

Different features of Flavobacteria, the most abundant class among the marine Bacteroidetes, have been studied in this thesis, but we particularly focused on two main traits: light sensing and polymer degradation. As a model member of Flavobacteria we chose *Polaribacter* sp. MED152. *Polaribacter* is one of the most abundant genera in the ocean. In 2004, 9% of the total number of marine Bacteroidetes 16S rRNA gene sequences deposited in GenBank belonged to the genus *Polaribacter* (Pommier *et al.*, 2005), and 36% of these were closely related to MED152. It has also been observed that MED152 is the tenth isolate that recruites more sequences in marine metagenomic studies (Yoosheph *et al*., 2010). Moreover, using FISH technique, *Polaribacter* has been found to account for 17% of the total cell

count in the Arctic Ocean (Nikrad *et al.,* 2012). Therefore, its high abundance in the marine bacterioplankton together with the possesion of the proteorhopsin gene, made MED152 a very attractive genome to analyze.

Thus, in **Chapter 1**, the study of the MED152 genome opened a window to the understanding of the evolutionary adaptation of these organisms to grow in oligotrophic sunlit environments. In order to check whether common patterns could be found within other Flavobacteria and compare them to other groups of bacteria, additional Flavobacteria genomes were analyzed in **Chapters 2** and **3**, including some representatives of the most abundant bacterial groups in the ocean like Proteobacteria and Cyanobacteria. Possible patterns due to the presence of the PR gene, in polymer degradation and in genomic islands content were examined. An attempt was made to check experimentally some of the hypotheses derived from these in silico studies in **Chapters 4** and **5** by means of pure cultures and biochemical tests in natural microbial assemblages. Figure 2 shows the relationship among chapters.

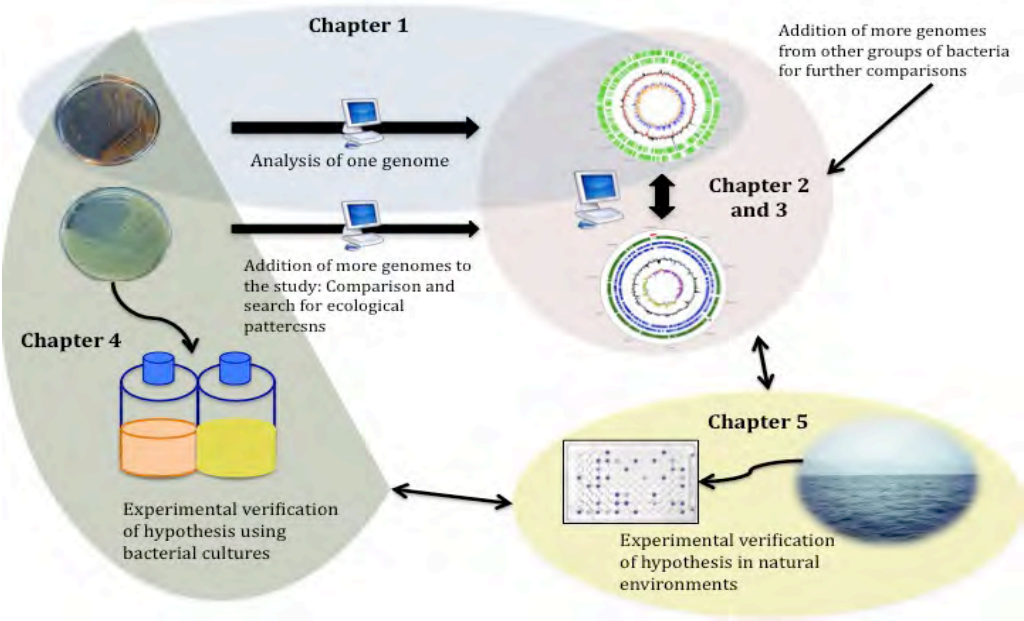


Figure 2. Relationship among the chapters of this dissertation. The genome of one Flavobacteria was analyzed in **Chapter 1**. In **Chapter 2** we compared four Flavobacteria searching for common patterns that could be present in their genomes and in **Chapter 3** we added even more genomes to the study of patterns in genomic islands in marine bacterioplankton. As some of these Flavobacteria had been isolated we could use them in **Chapter 4** to do *in vivo* experiments trying to test out some of the questions that appeared form the study of the genomes. Finally, in **Chapter 5** we went directly to the environment and studied the microbial community *in situ.*

Alternation of lifestyles to survive on scarcity

It is risky to discuss the ecology of an organism only by analyzing its genome, but the genes found in a genome can provide a lot of information about its potential strategies. In **Chapter 1,** the genome of *Polaribacter* sp. MED152 suggested a potential alternation of lifestyles according to the levels of organic matter in the water. If high-molecular-weight particles (HMW-DOM) are abundant, MED152 could colonize them and move around them by gliding motility. But once all the polymeric substrates had been consumed, MED152 would have to carry a free-living existence thanks to its ability to respond to light. Moreover, an increase in CO_2 fixation in the light suggested that the limited central metabolism could be complemented by anaplerotic inorganic carbon fixation.

a) **Living in abundance: Adhesion and degradation of polymers**

Generally, Bacteroidetes are considered to be efficient consumers of polymers such as polysaccharides and proteins (Cottrell and Kirchman, 2000; Woyke *et al*., 2009). Thus we should find genome evidences that would corroborate that fact in *Polaribacter*'s. In effect, MED152 had a large glycolytic potential, including 29 glycoside hydrolases and 81 peptidases. It is also possible that MED152 could produce exopolysaccharides involved in non-specific attachment to particles, since its genome encodes 32 glycosyl transferases (Corpe *et al*., 1976; Sutherland *et al*., 1985).

It is also known that polymeric organic matter can be bound by outer membrane complexes (Fig. 3), formed by SusC, a ligand-gated channel, and SusD, an outer membrane protein (SusCD complexes), which function as polysaccharide binding entities before starting its degradation (Anderson and Salyers, 1989; Blanvillain *et al*., 2007; Shipman *et al*., 2000). In these complexes, hydrolytic enzymes are located on the surface of the bacterium, but there are also outer membrane proteins that bind the polysaccharides to the bacterial surface (SusD) (Reeves *et al*., 1997). Five SusCD complexes were detected in MED152 (**Chapter 1**, Fig. S7; **Chapter 2**, Table SM9). This is a clear evidence that adhesion and degradation are tightly coupled for efficient utilization of polymers. Recent studies have found that TonB-dependent transporters were very abundant in marine microbial communities and even more, the energy to transport solutes could be provided by the PR (Morris *et al*., 2010). In fact, TonB-dependent transporters are, together with ABC transporters, the most abundant type transporters in MED152, representing 3.9% of the total amount of transporters.

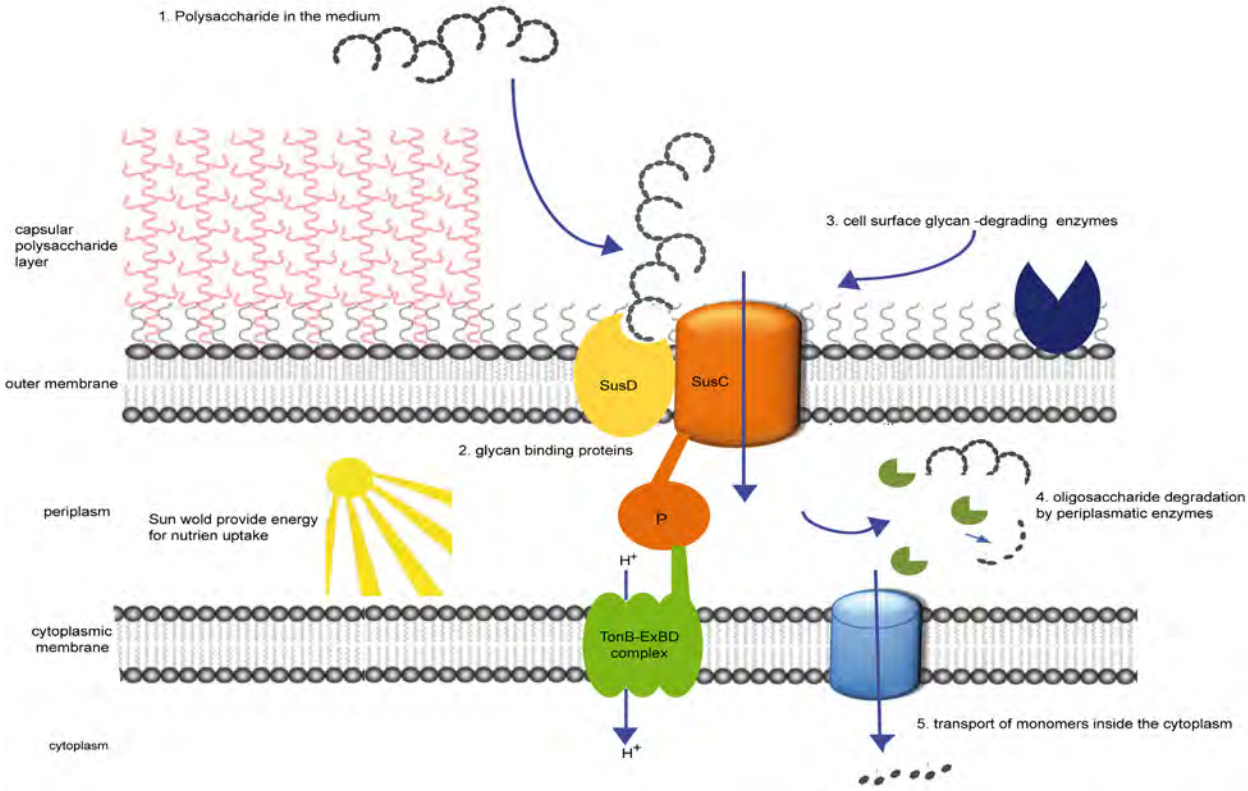


Figure 3. Simplified model for polysaccharide processing based on Sus System in *Bacteroides thetaiotaomicron*. Adapted from Martens EC *et al*., J. Biol Chem (2009).

Moreover, MED152 had a complete set of genes (15) involved in gliding motility, which would be beneficial in the exploration of solid surfaces. Although it is not totally understood, it seems that the ability to glide over surfaces may provide cells with a selective advantage for utilization of polymers. This is how it would work the predicted mechanism of *C. hutchinsonii* (Xie *et al*., 2007) and *F. johnsoniae* (Braun *et al*., 2005) for cellulose and chitin utilization respectively.

Furthermore, up to 26 ORFs contained domains that are involved in the attachment to surfaces (**Chapter 1**, Table SM9), some of them within hydrolytic clusters that included SusC and SusD (**Chapter 1**, Fig. SM7).

b) **When DOM is depleted: light-harvesting system**

Light is an important environmental factor and the principal source of energy in the biosphere. Thus, organisms have evolved means to sense and respond to light in order to optimize their growth and metabolism. One of the key points in the genome of MED152 is

the presence of the proteorhodopsin (PR) gene and the genes for the synthesis of retinal (PR chromophore). Besides it harbors many genes involved in light regulation responses (**Chapter 1**, Table SM6). It was very interesting to find out that MED152 had a phytochrome, a well known protein that responds to the red or far-red regions of the visible light spectrum and is also present in Cyanobacteria and plants, where phytochrome activation regulates approximately 10% of plant genes (Sharrok, 2008). Although the complete signal transduction pathway for a plant phytochrome response has not yet been described, there is evidence that plant phytochromes have signal transduction activity (Sharrok, 2008). Thus, its presence in MED152 suggests that this protein could contribute to the regulation of the synthesis of proteins that respond to light.

PR-phototrophy had been suggested to supplement the energy requirements of the cell in the presence of light, decreasing the amount of carbon respired. However, this represents a challenge for the cell, resulting in an imbalance of the TCA cycle (Krebs Cycle), that MED152 could solve fixing CO_2 by anaplerotic reactions, generating oxaloacetate from bicarbonate by means of PEP and pyruvate carboxylase enzymes (**Chapter 1**, Fig. SM4). When light is the sole source of energy (ATP), the TCA cycle is likely used for biosynthetic purposes and not for energy generation (as in autotrophic organisms). The glyoxilate shunt (also present in MED152, **Chapter 1**, Fig. SM3) allows growth on some compounds that enter the right hand part of the TCA cycle, such as lipids or acetate. But in MED152 that can obtain energy from light this metabolic pathway could play an important role regenerating oxaloacetate. The glyoxylate shunt is not present in autotrophic bacteria, but it is in plants and many other bacteria.

Figure 4 compares *Prochlorococcus marinus* and *Polaribacter* sp. MED152. Notice the horseshoe type of TCA cycle in *Prochlorococcu*s. This is typical of obligate autotrophs because they don't need to complete the cycle since it is not oxidative. Thus they lack one of the enzymes: the alpha-ketoglutarate dehydrogenase.

Interestingly, MED152 fixed more bicarbonate in the light than in the dark when growing in a relatively poor medium (8-fold diluted Marine Broth) (**Chapter 1**, Fig. 3 and **Chapter 4**, Fig. 7). This suggested that light-stimulated anaplerotic inorganic carbon fixation could be a means allowing proteorhodopsin-containing Flavobacteria to efficiently use organic matter for biosynthesis.

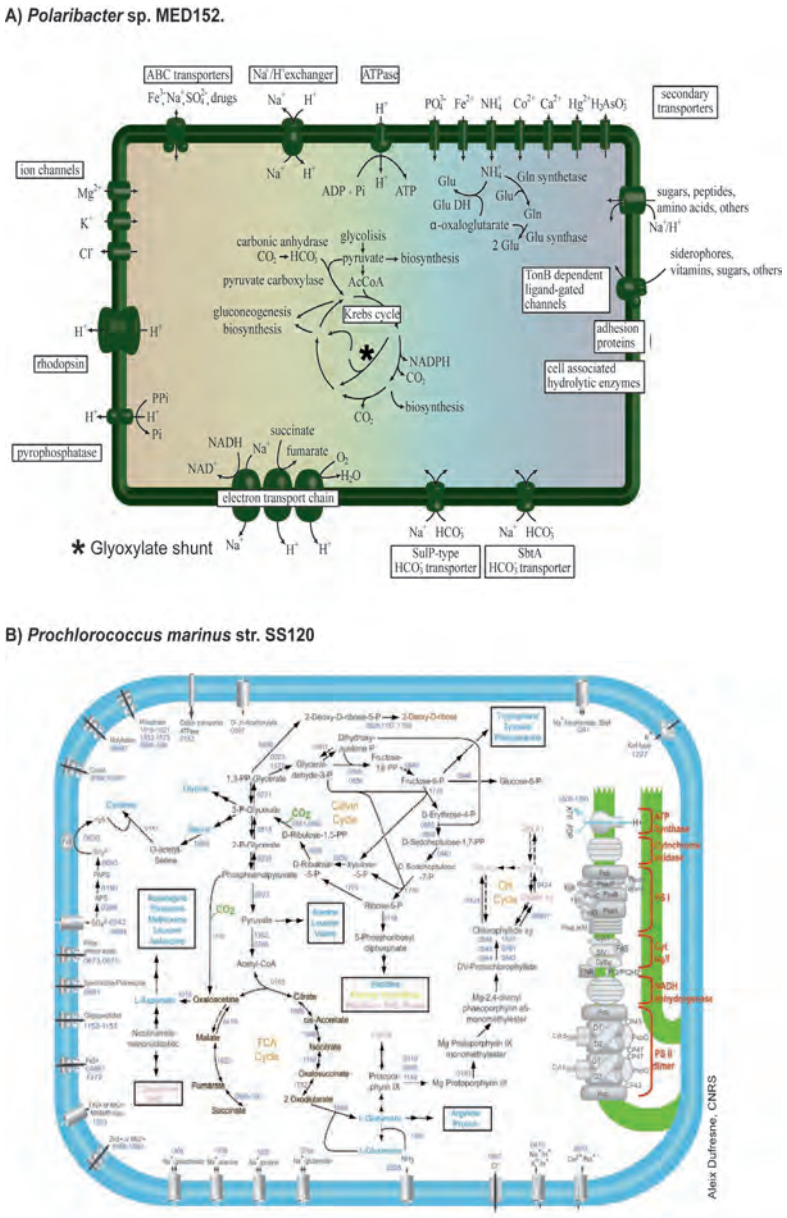


Figure 4. Comparison between (A) *Polaribacter* sp. MED152 and (B) *Prochlorococcus marinus* str. SS120. Transporters are similar because they take up the same nutrients and the also rely on the Na^+ gradient. Some important differences are the Calvin cycle and the right hand part of the cell where the photosystems and electron chain appear. MED152 does not have the Calvin cycle or any other cycle to fix CO_2 autrotrophically.

In summary, having the ability to attach to particles, degrade them, and incorporate the resulting small molecules inside the cell, MED152 could survive while there is particulate organic matter present in the ocean. Once these particles were exhausted, the PR would provide the energy necessary to survive without requiring organic compounds to respire. Figure 5 shows a hypothetical model for the strategies of MED152 in the ocean.

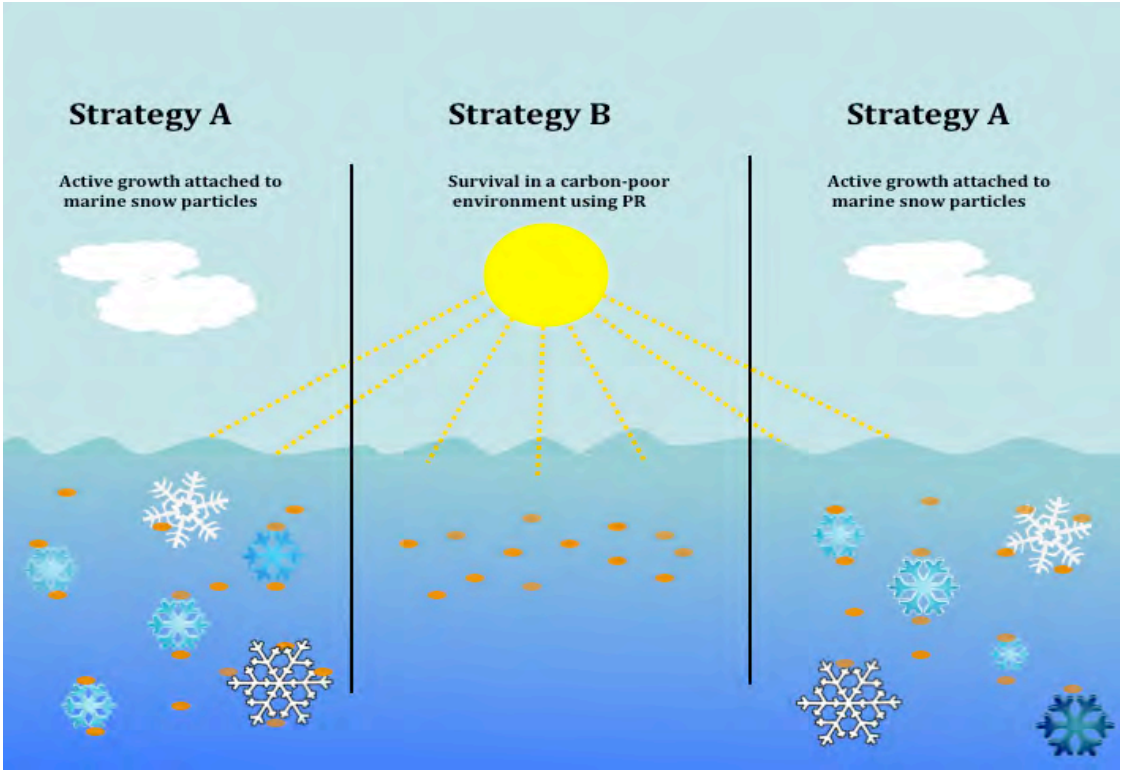


Figure 5. "Traverse of the Desert Model" for *Polaribacter* sp. MED152 in the Ocean.

Patterns among marine Bacteroidetes

The first whole marine Bacteroidetes genome analyzed was that of *Gramella forsetii* KT0803 (without PR)(Bauer *et al*., 2006), indicating a clear adaptation to the degradation of HMW-DOM compounds. *Polaribacter* sp. MED152 was the second analyzed genome of this class showing the same adaptations for degrading polymers (**Chapter 1**). Given that this ability of degradation appeared to be a recurring pattern and independent of the fact of harboring the PR gene, the comparison with more marine Bacteroidetes genomes, with (PR+) and without (PR-) proteorhodopsin, allowed us to identify the general patterns among representatives of this taxonomic class.

a) **Correlation between possession of the PR gene and genome size.**

It is intriguing that all the Bacteroidetes PR-containing genomes are small (**Chapter 2**, Fig. 1). As seen in **Chapter 1**, the genome of MED152 is very small (2.9 Mb). MED134 also carries the PR gene and both of them have genomes smaller than the other two genomes without PR, *Gramella forsetii* and MED217 (**Chapter 2**, Table 1). It is tempting to think that

somehow the PR allows bacteria to reduce their genomes, so that they may not need to carry the genes for many different carbon utilization pathways because light energy would make them more independent from organic compounds as energy sources.

b) **Are PR+ bacteria better adapted to live under oligotrophic conditions?**

In **Chapter 1** we proposed that PR, together with other light sensing genes and the anaplerotic carbon fixation, represented adaptations to survive in the surface ocean when concentrations of OM were low. The next step was to test whether light sensing capacity and the existence of CO_2 fixation in other Flavobacteria were features consistent with the presence or absence of PR. Concerning light sensing domains, no significant differences were observed in number comparing PR+ and PR- bacteria (**Chapter 2**, Fig. 5). Nevertheless, incorporating more groups to the analysis such as Cyanobacteria or Proteobacteria, different patterns in light sensing domains came out (**Chapter 2**, Table 3), suggesting a differential light use strategy. With respect to the enzymes involved in CO_2 fixation (discussed in **Chapter 1**: pyruvate and PEP carboxylase, isocitrate lyase, malate synthase), there was a clear correlation between PR-possession and the number of the formerly mentioned genes (**Chapter 3**, Fig. 6B).

c) Do PR+ bacteria follow the same strategy as MED152 when they are short of organic matter?

The adaptations to particle attachment and degradation of MED152 were already mentioned in **Chapter 1**. We observed that all four Flavobacteria had large number of proteins with surface adhesion domains (**Chapter 2**, Table SM1). That was also true for other Flavobacteria, however, this feature did not appear in other groups of marine bacteria (**Chapter 2**, Fig. 8A and 8B). This seems to be a common trait among flavobacterial genomes independently to the ability to respond to light, since no differences between PR+ and PR- species were observed.

The degradative potential seen in **Chapter 1** was also corroborated with additional Flavobacteria (**Chapter 2**, Fig. 8C and 8D), which in general (with some exceptions) presented the largest amount of glycoside hydrolases and peptidases per megabase compared with other groups. Besides, PR+ Bacteroidetes showed significantly more peptidases per megabase that PR- Bacteroidetes, suggesting specialization for protein degradation (**Chapter 2**, Fig. 10) and they all showed different suits of these enzymes, indicating a specialization for particular types of proteins (**Chapter 2**, Fig. 11). Thus, each marine Flavobacteria is adapted to slightly

different niches thus decreasing competition.

Figure 6 plots the number of glycoside hydrolases, peptidases, glycosyl transferases, and adhesion proteins in groups such as Bacteroidetes, Planctomycetes, and Proteobacteria. Notice that our four Flavobacteria have the highest number of glycoside hydrolases (GHs) per megabase (Mb) after the two *Bacteroides* spp. which are well-known specialists for polysaccharide degradation, and they harbor the highest number of peptidases per Mb after the gammaproteobacterium *Idiomarina loihiensis* which has been proposed to rely mainly on amino acids for carbon and energy supply (Hou *et al*., 2004).

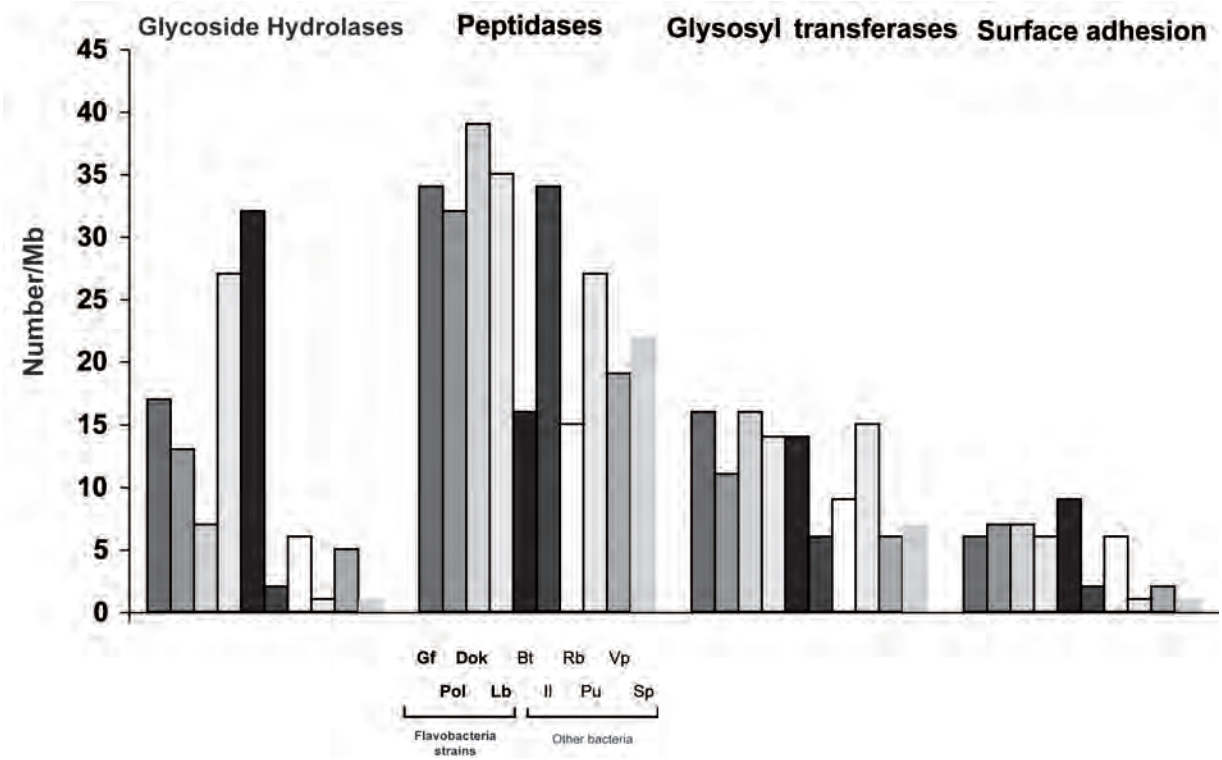


Figure 6. Comparison of hydrolytic capabilities and adhesion potential of marine heterotrophic bacterioplankton and human-associated members of the Bacteroidetes. **Gf:** *Gramella forsetii*, **Dok:** *Dokdonia* sp. MED134, **Pol:** *Polaribacter* sp. MED152, **Bt:** *Bacteroides thetaiotaomicron*, **Lb:** *Leeuwenhoekiella blandensis* MED217, **Rb:** *Rhodopirellula baltica*, **Il:** *Idiomarina lohiensis*, **Vp:** *Vibrio parahaemolyticus*, **Pu:** *Pelagibacter ubique*, **Sp:** *Silicibacter pomeroy*i.

Quantification of Genomic Islands (GIs) in marine bacterial

Bacterioplankton live and compete in the ocean's photic zone under the pervasive influence of light (Frigaard *et al*., 2006). The only way of avoiding competition is evolving different life strategies. Horizontal gene transfer is one of the evolutionary mechanisms enlarging the flexible genome pool of bacterial populations, facilitating their adaptation to new ecological niches (Doolittle 1999; Boucher *et al*., 2003). The new genes transferred

are clustered in what is called genomic islands (GIs). Recently, it has been observed that *Dokdonia* sp. MED134 has the genes directly involved in PR phototrophy (*prd*:PR and *blh*: ß-carotene) and a blue light sensor close to a conjugative transposon (González *et al*., 2011). These genes could be transferred to other genomes and if they harbor the genes for the synthesis of carotenoids, they easily become phototrophs. Likewise, many other genes could provide a potential benefit for increasing bacterial fitness if transferred laterally.

This was the next step in the study of the genetic potential o marine Bacteroidetes. For this purpose, in **Chapter 3** we constructed a database that included 8152 genes present in GIs of 66 genomes out of the 70 chosen bacterial genomes. In **Chapter 2**, we saw that GIs were one of the features that distinguish one Flavobacteria from the other by analyzing one GI of *L. blandensis* MED217 (**Chapter 2**, Fig. 2), but this time we studied more flavobacterial genomes and we added genomes from other bacterial groups in order to compare them. On the whole, 438 GIs were detected, spanning a total of 8.87 Mb (**Chapter 3**, Table SM2). Unfortunately, it was impossible to detect genomic islands manually accurated as it had been done for some *Prochlorococcus* (Coleman *et al*., 2006) and *Synechococcus* (Dufresne *et al*., 2008) strains due to the large number of genomes and because some of them were not closed yet. For this reason we had to resort to an automatic search tool and we ascertain that although missing some GI we could be sure that what we detected was a real GI (**Chapter 3**, Fig. 1, Table 1, and Table SM1). GIs accounted for 0-12% of the genomes and most of the genomes ranged between 2 and 5 % (**Chapter 3**, Fig. 3).

Architecture and functional annotation of marine bacterial GIs

If horizontal gene transfer (transformation, conjugation and transduction) is a mechanism by which bacteria acquire new genes, one would expect to find a large number of mobile genetic elements such as transposases, conjugative transposons, integrases or phage integrase-related genes (Dobrindt *et al*., 2004, Juhas *et al*., 2008). In effect, we found that 70% of the described GIs contained these mobile genetic elements.

As expected, we found that most GIs had been originated through horizontal gene transfer (**Chapter 3**, Fig. 4), by presenting high percentage of hypothetical proteins and different set of genes in different genomes. Interestingly, we found another structure of GI, transferred via homologous recombination (**Chapter 3**, Fig. 5). This kind of GI is characterized by having housekeeping genes, low number of hypothetical proteins and the same structure in different organisms. Flavobacteria was one of the groups with higher numbers of this type of islands.

To demonstrate the crucial role of GIs in the evolution of bacteria, it is necessary to check the function of the genes within the GIs. Around 55-60% of the genes could not be assigned to any functional category, therefore, these hypothetical proteins could be considered as novel. The remaining genes within the GIs were annotated in order to assign them to a functional category using two approaches (**Chapter 3**, Fig. 8 and Fig. 9), revealing a diverse range of biologically relevant genes were present in the GIs.

Furthermore, it is obvious that different groups of marine bacteria have different strategies for living in the ocean. The fact of finding differences in the functional categories represented in our 4 groups would reveal one reason for its adaptation to the marine environment. Thus, the comparison of the genes with functional categories assigned (**Chapter 3**, Table 2 and Table SM4) showed significant differences in all cases, suggesting that each phylogenetic group could have a different strategy in the role played by GIs in their ecology. It is probable that many genes within GIs are positively selected due to the potential benefits of these genes for the bacterial life-style. They may benefit bacterial fitness under environmental changing conditions by acquiring novel foreign genes.

Beyond the genomes: analizing the physiological function of PR

Since the PR gene was discovered (Béjà *et al*., 2000), its presence in several groups of marine bacteria has been revealed, although a consistent physiological function for PR has not emerged yet. Three chapters of this thesis (**Chapters 1, 2** and **4**) focused on the implications/benefits of possessing the PR. **Chapter 1** and **2** tried to hypothesize its environmental role from what "is written" on the bacterial genomes. **Chapter 4** tried to analyze the physiological function of this protein by means of growing experiments with two PR-carrying isolates.
As exemplified in Figure 7, seawater culture experiments with organisms that have been previously isolated are necessary to understand the physiology and the potential importance in the environment of the PR. Experiments with artificial seawater (ASW) showed different growth responses under several experimental conditions.

In **Chapter 1** we hypothesized the possible function of PR in MED152. Based on some of the results obtained in **Chapter 2**, we already mentioned that there were features in the content of the different genomes that suggested different ways of using the PR. The results obtained in **Chapter 4** revealed a slightly different use o fPR between MED152 and MED134,

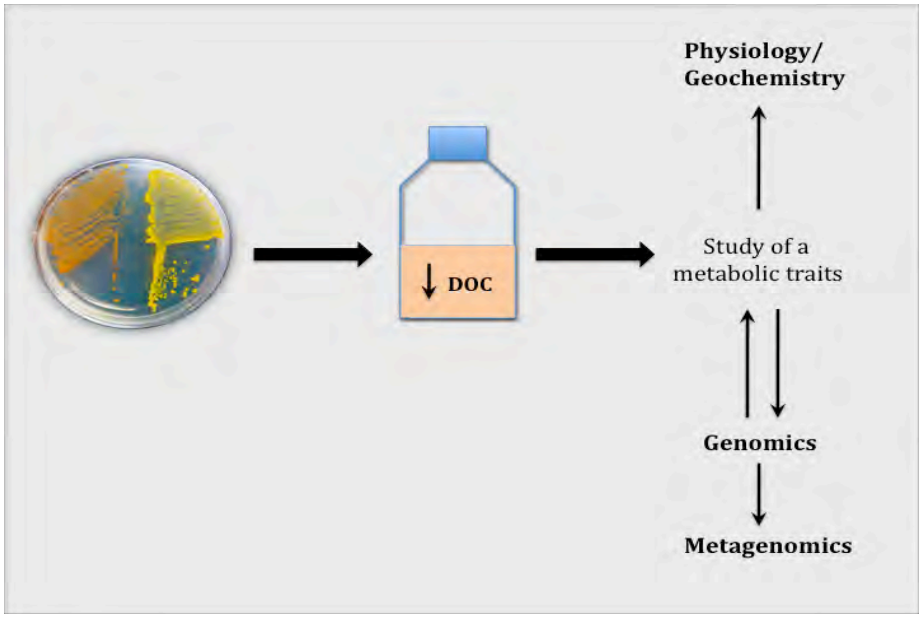


Figure 7. Schematic representation of the importance of pure cultures for bacterial function determination.

likely in order to adapt to slightly different niches. While MED34 showed higher differences in yield between light and dark incubations (Gómez-Consarnau *et al*., 2007) than MED152 (**Chapter 4**, Fig. 6); MED152 showed greater rates of carbon fixation (**Chapter 4**, Fig. 7A). It was intriguing to observe that the differences in response to light, besides occurring at low organic matter concentrations, also depended on the composition of this organic matter (**Chapter 4**, Fig. 4-6). Unfortunately, our results do not clarify this observation and, therefore, more experiments must be done in order to demonstrate the actual effects of PR with different media and growth conditions.

Arctic microbial community metabolic capacities

Coming back to the hydrolytic potential of marine Bacteroidetes discussed in **Chapters 1** and **2**, in **Chapter 5**, polymer (and other substrates) degradation capacity was assayed in the field using a cruise of opportunity during the International Polar Year (IPY) using Biolog plates as a metabolic fingerprinting technique.

It had already been seen that Biolog Ecoplates were useful to track the metabolic capacities of the microbial assemblages in a previous Arctic cruise (2003-2004). In **Chapter 5** we developed a customized Biolog plate, the so-called MT2 plate, with the intention of analyzing the degradation of several compounds by the natural community (**Chapter 5**, Table 3) at different depths and along the year (February to July).

Unlike Ecoplates, MT2 plates have to be complemented with the different carbon sources to be studied, but the first problem that arises is to determine the best concentration for each of the compounds to be tested in order to detect oxidation. This was difficult to decide since the concentration of the substrates in the Ecoplate (some of which we added to the MT2), our reference, are not facilitated by the manufacturer. Also, not all substances behave in the same manner at certain concentrations, while some can be water soluble, others (especially polymers) can precipitate. Moreover, A the study of functional diversity measured as a carbon source utilization using Biolog MT2 plates had been conducted mostly in microbial communities from soils, in some cases from fresh water, and just in one occasion was used to determine the potential utilization of compounds by bacteria in marine waters. Thus, from the 23 substrates tested, only 17 were used (**Chapter 5**, Fig. 2, Table 2), but we do not rule out that the rest of compounds were not used, simply that there was no oxidation observed, maybe because the bacterial activity was not enough, maybe because the concentration of the compound was too low.

We observed that bacteria at the surface utilized more variety of substrates than at depth (**Chapter 5**, Fig. 2 and Fig. 3) and more frequently, although these differences were not significant for any of the compounds except for the whole set of amino acids taken together. With respect to the time of the year, there was no detectable activity during winter, but the number of positives started to raise with the increase in chlorophyll *a* in the water, from mid-March onwards, except for a period of 4 weeks were the signal disappeared completely (**Chapter 5**, Fig. 3). We hypothesized that this "hole" found in the middle of a highly activity period could be due to a change in the bacterioplankton community (observed in a DGGE gel) due to the transition from winter to spring conditions, when the winter-adapted assemblage was disappearing and the spring/summer-adapted assemblage was appearing (**Chapter 5**, Fig. 1SM).

The results were compared with those from a previous study (CASES) carried on in the same region but in a different hydrographical area, Franklin Bay. In the latter study, polymeric carbohydrates produced the largest number of positive results, like in CFL, but amino acid showed a much less proportion of positive results (**Chapter 5**, Fig. 5). Figure 8 shows the difference in the most used compound during CFL (A) and CASES (B). The proportion of polymers and carbohydrates used were higher in CASES and the proportion of amino acids was higher in CFL. In both studies the activity at the surface was higher than at depth.

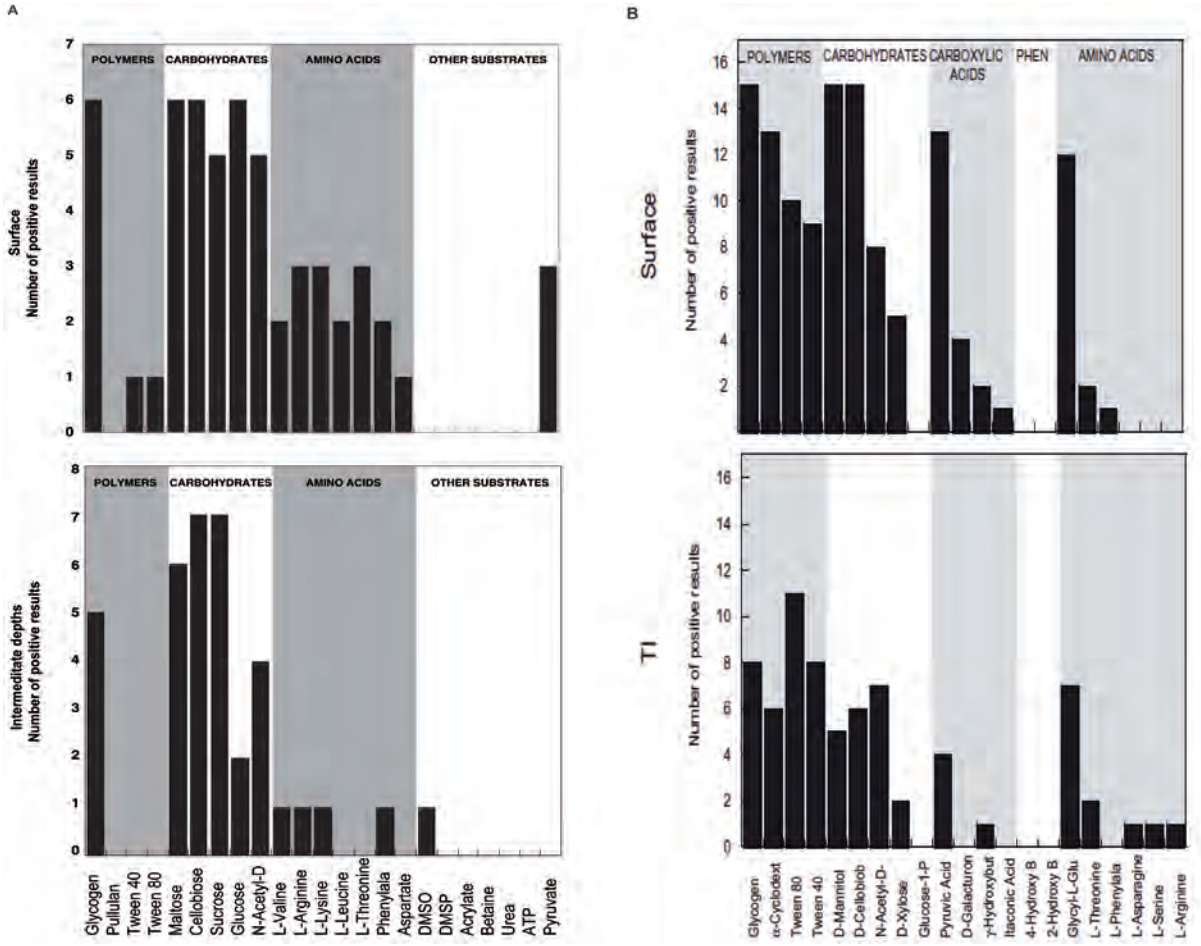


Fig.8. Number of positives results for substrates in the Biolog MT2 microplates during CFL (A) and in the Biolog Ecoplate during CASES (B) studies.

Another difference with CASES was the time of the year with more activity. While the most active part of the year in CFL was from spring onwards, following the same pattern as chlorophyll *a*, in CASES, winter was the most active period of time and the activity decreased when spring started (Figure 9). It is difficult to determine to the cause of these differences between the two studies, but one likely reason could be the different hydrographic conditions of Franklin Bay with respect to Amundsen Gulf.

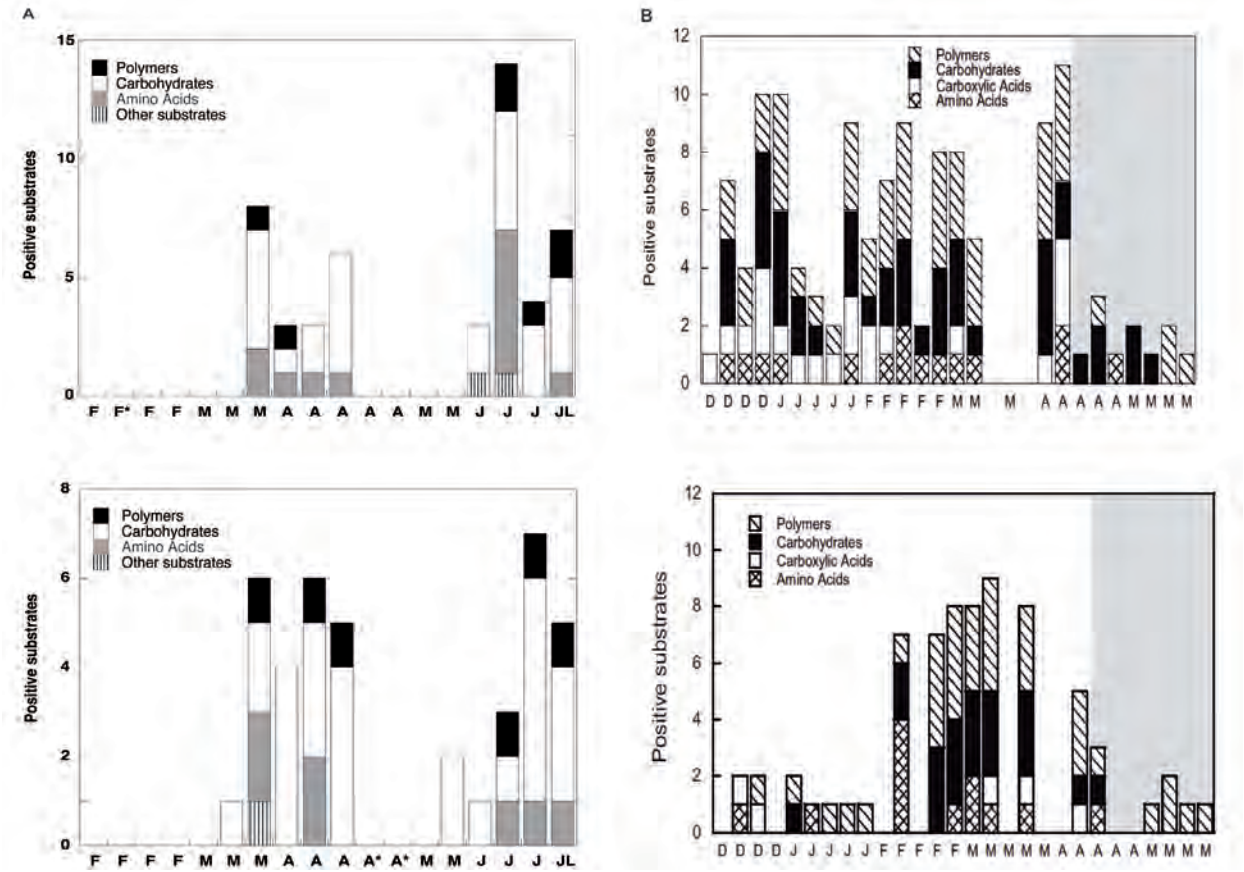


Figure 9. Number of substrates used along the year, and separated into categories for CFL (A) and CASES (B).

CONCLUSIONS

1. The general features of the *Polaribacter* sp. MED152 genome are consistent with life in the surface ocean, including light utilization and a significant number of Na^+ dependent proteins.

2. More specific features of the *Polaribacter* sp. MED152 revealed adaptations to alternate between attachment to particles and free-living life styles.

3. Using proteorhodopsin, *Polaribacter* sp. MED152 is able to capture light energy, increase anaplerotic bicarbonate fixation and likely, therefore, it will use the few carbon compounds it can get hold of exclusively for biosynthesis.

4. Marine Bacteroidetes share the capacity for adhesion, gliding motility, and a large number of polymer degrading enzymes.

5. Marine Bacteroidetes containing PR have especially small genomes with more genes involved in CO_2 fixation per Mb than the Bacteroidetes without PR.

6. Genomic islands are a common feature in most of the marine bacterial genomes examined. Each bacterial class favors a different set of genes, indicating different strategies in the role played by GIs in their ecology.

7. Horizontal Gene Transfer and Homologous Recombination have an important role for the mobility of DNA clusters of genes between distinct taxa and within closely related genomes respectively modulating the flexible pool of the genome.

8. Functional annotation of the genes within the genomic islands revealed the presence a diverse range of biologically relevant genes.

9. Light stimulates growth of the proteorhodopsin-containing Flavobacteria *Polaribacter* sp. MED152 in limited organic matter environments.

10. Culture medium composition influences whether differences are found in growth between light and dark conditions.

11. Proteorhodopsin enhances bicarbonate uptake in *Polaribacter* sp. MED152 growing in different poor media.

12. Carbohydrates, particularly polymeric molecules, are an important carbon source for the Arctic bacterioplankton. Amino acids were also consumed but their uptake is apparently rather variable depending on the areas and the season.

FUTURE PERSPECTIVES

The interest in Flavobacteria is increasing every day, together with the number of microoorganisms of this group that are being isolated in pure culture and the genomes sequenced. Based on the results obtained in this dissertation, the consideration of some of the following aspects would increase understanding of the role of the microorganisms (in our case the marine Bacteroidetes) in the ocean:

1. Complete sequencing and annotation of more members of marine and non-marine Bacteroidetes. Although the number of Bacteroidetes isolates is high, to study them it would be necessary to have their entire genomes and all annotated genes in order to understand and compare them with other representatives of the same or different groups.

2. From all the genomes that have been analyzed, a big effort should be done in order to try to discover the functions of the hypothetical proteins, which always represent a high percentage of the genome and, probably, these proteins contain the essence that makes each organism unique.

3. Improvement and development of bioinformatics techniques for analyzing and comparing many genomes at once. Every day, there is more and more information to process and every time, more sophisticated and powerful machines would be necessary to cope with thousands of gigabytes of data. Besides, new software that allows efficient analysis in a relatively short time would be required.

4. More effort is needed to develop a completely defined culture medium for *Polaribacter* and *Dokdonia*. Likewise, more studies are necessary to determine why light makes a difference only with certain carbon sources and not others.

5. Finally, we should design new BIOLOG plates, more specific and trying new substrates not used so far and that could increase our understanding of metabolic diversity heterotrophic bacterioplankton.

Resumen en Español (Spanish Summary)

INTRODUCCIÓN GENERAL

EL AMBIENTE MARINO

Los océanos son el mayor reservorio de agua en el Planeta. Con casi $1.37x10^9$ km^3 y aproximadamente $3.61x10^{14}$ m^2, los océanos cubren alrededor del 71% de la superficie de la Tierra (Davis Jr., 1972). Más de la mitad de este área tiene una profundidad de más de 3.000 metros, siendo la Fosa de las Marianas el punto más profundo, con alrededor de 11.000 metros de profundidad. La salinidad del agua de mar está entre las 33 y 37 con un promedio 35. El pH en superficie es de alrededor 8.1 de media, acidificándose progresivamente desde que comenzó la revolución industrial (Orr *et al*., 2005). La temperatura media de las aguas superficiales del océano es de unos 17°C, cambiando sustancialmente con la latitud. En los polos la temperatura puede ser de -2°C, mientras que en el Golfo Pérsico puede llegar a 36°C.

Los océanos influyen significativamente sobre el clima de la Tierra (Stewart, 2008) mediante la redistribución de calor desde los trópicos hasta las regiones polares (circulación impulsada por el viento y la circulación termohalina), intercambiando calor y vapor de agua con la atmósfera (que influyen en los patrones de precipitaciones y la circulación atmosférica) y actuando como fuentes y sumideros de CO_2. Se cree que las perturbaciones en la circulación termohalina podrían tener un impacto considerable en el clima de la Tierra (Bryden *et al*., 2005).

El medio marino contiene una gran cantidad de carbono. Una parte está en forma de carbono orgánico, que puede estar tanto en forma disuelta (carbono orgánico disuelto, DOC) como particulada (carbono orgánico particulado, POC). El carbono impulsa la vida en los océanos, ya que es fijado por el fitoplancton a través de la fotosíntesis, proporcionando la base de las cadenas alimentarias marinas. A pesar de hallarse en concentraciones relativamente bajas, nutrientes tales como el nitrógeno (N) y el fósforo (P) son particularmente importantes para la biología de los océanos. Estos nutrientes son limitantes en la mayoría de las regiones productivas. En algunos casos, el hierro (Fe) y el sílice (Si) también pueden actuar como nutrientes limitantes (sobre todo en las regiones de mar abierto). La relación entre la concentración de carbono, nitrógeno y fósforo en el fitoplancton es 106:16:1, la clásica "relación de Redfield" (Redfield, 1934).

EL BACTERIOPLANCTON Y SU PAPEL EN EL OCÉANO

La vida en los océanos se originó al menos 3 mil millones años antes que en la tierra. Los microorganismos han sido las únicas formas de vida que han estado presentes el 80% de la historia del planeta y toda la vida multicelular depende de procesos microbianos (Amaral-Zettler *et al.*, 2010). Por bacterioplancton se entiende el conjunto de bacterias y arqueas del plancton (Thurman, 1997) y son los organismos más pequeños y más

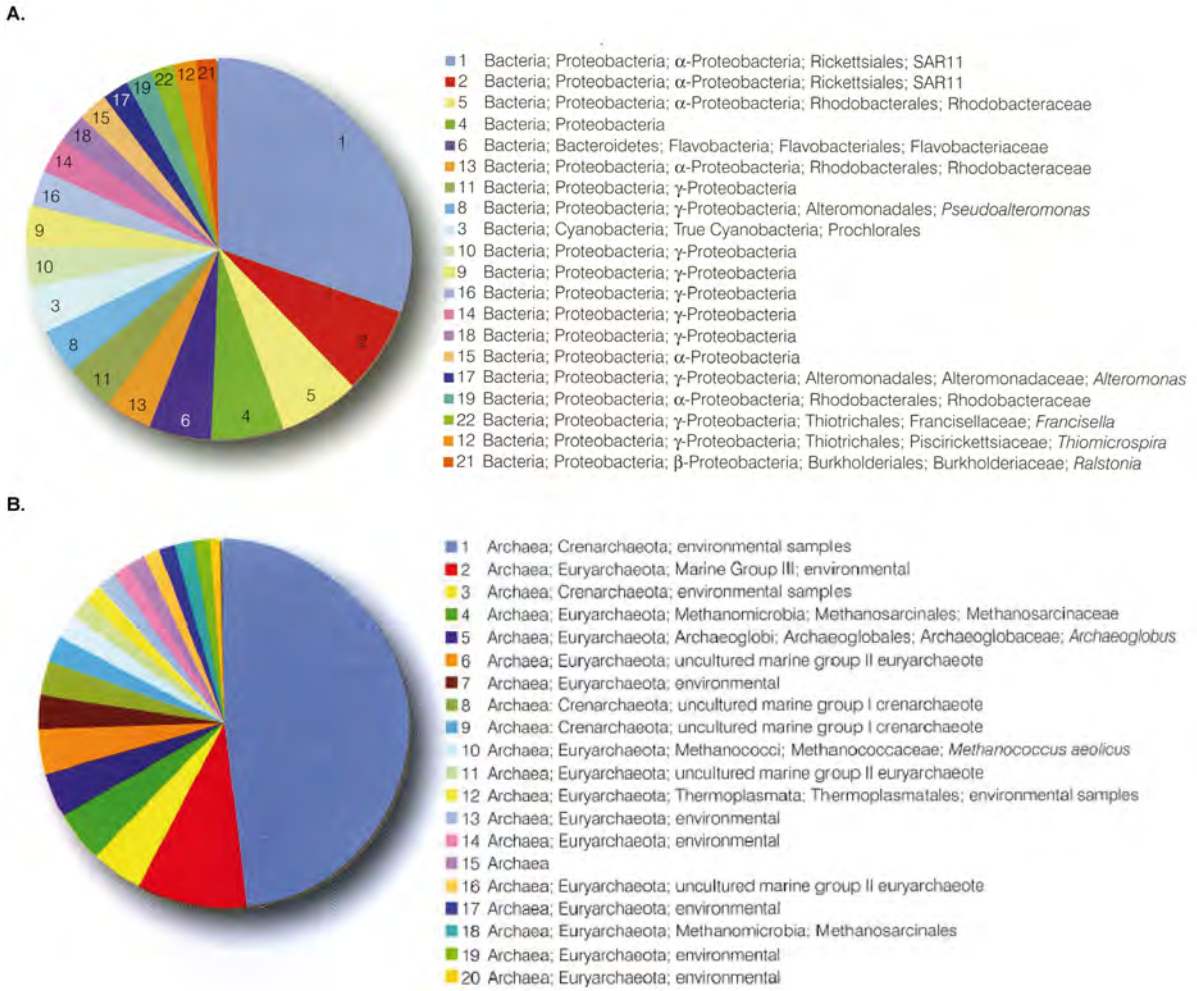


Figura 1. (A) Desglose taxonómico de las 20 secuencias bacterianas más abundantes en 583 bases de datos bacterianas. (B) Desglose taxonómico de las 20 secuencias de arqueas más abundantes en 120 bases de datos bacterianas. El ranking de ambos gráficos se basa en la suma de las abundancias relativas de secuencias individuales de cada muestra. Las taxonomías están en base al procedimiento *Global Alignment for Sequencing Taxonomy* (GAST) (Huse *et al.*, 1997). Figura procedente del libro A *Global Census of Marine Microbes* (Amaral-Zettler *et al.*, 2010).

abundantes en el océano (Whitman *et al*., 1998). La función ecológica de las bacterias y arqueas marinas es ciertamente importante para la vida en la Tierra ya que controlan la ecología del ambiente marino por su papel como principales productores y consumidores de la materia orgánica disuelta (DOM) (Kirchman, 2008). Son las formas de vida más antiguas, los catalizadores de la transformación de energía primaria y son fundamentales para los ciclos biogeoquímicos que modelan nuestro planeta (Falkowski *et al*., 2008). Son muy abundantes, con concentraciones alrededor de 10^5 células por cm^3 (Hobbie *et al*., 1977; Zimmerman R. and Meyer-Reil, 1974) y extrapolando a todo el océano, considerando un volumen de $1.4x10^9$ km^3, el número total asciende a aproximadamente 10^{29} células procariotas (Whitman *et al*., 1998). Para ponerlo en perspectiva, se puede considerar que sólo hay 10^{11} neuronas en un cerebro humano o estrellas en nuestra galaxia.

Los océanos albergan numerosos y diversos grupos de bacterias y arqueas. Muchos de ellos son autótrofos, que comprenden el fitoplancton, y sintetizan materia orgánica a través de la fotosíntesis. Los productores primarios en el océano son responsables de cerca de la mitad de la producción primaria total del Planeta (Falkowski *et al*., 1998; Field *et al*., 1998). Comprenden el grupo de las Cianobacterias, que incluye los dos géneros más abundantes en el océano: *Prochlorococcus* (Chisholm *et al*., 1988) y *Synechococcus* (Waterbury *et al*., 1979). Sin embargo, la mayor parte de las bacterias marinas son heterótrofas y adquieren la energía mediante el consumo de materia orgánica. Ésta es producida mayormente por el fitoplancton (Fogg, 1958; Sharp, 1977), aunque los herbívoros (Nagata, 2000) y los virus, mediante la inducción de la lisis celular (Wilhelm y Suttle, 1999), también contribuyen al conjunto de la materia orgánica. Las bacterias heterotróficas oxidan aproximadamente la mitad del carbono labil fijado por la fotosíntesis (Azam *et al*., 1983b; Fenchel, 1988; Robinson, 2008), aunque esta proporción puede ser mucho mayor en aguas oligotróficas (del Giorgio et al., 1997). Como consecuencia, el dióxido de carbono es liberado al océano y una parte de los componentes orgánicos se remineraliza liberando nutrientes inorgánicos esenciales. Esta biomasa puede ser reintroducida de nuevo en la cadena trófica y representa una fuente adicional de sustratos orgánicos para los niveles tróficos superiores (protistas), limitando al mismo, tiempo la pérdida de materia orgánica a través de la sedimentación y enterramiento profundo. Por lo tanto, las bacterias desempeñan un papel crucial en las redes tróficas pelágicas mediante el control de gran parte de los flujos de carbono en el océano (Azam, 1998).

Esta cadena alimentaria en particular es conocida como el "bucle microbiano" (Azam *et al*., 1983a; Ducklow, 1983; Pomeroy, 1974). En la actualidad, con el aumento de los niveles de CO_2 en la atmósfera, comprender la función de los microorganismos en las redes alimentarias marinas es un tema esencial ya que el bucle microbiano determina, en parte, la respuesta de los ecosistemas oceánicos y el ciclo del carbono al cambio climático (Falkowski *et al*., 2000; Kirchman *et al*., 2009).

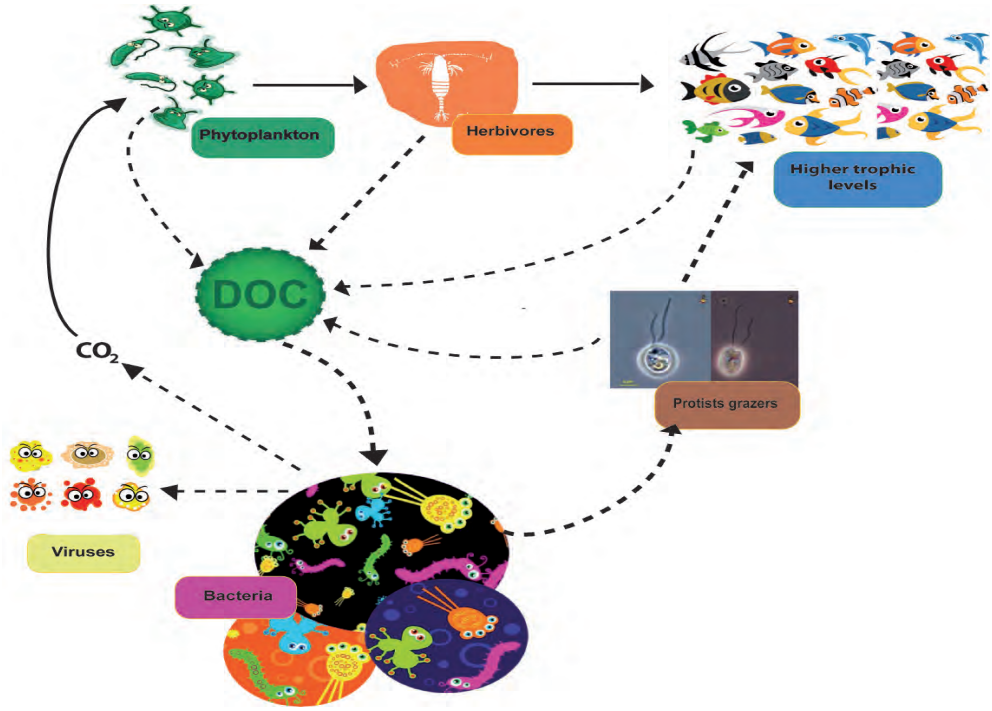


Figura 2. Diagrama de la red trófica microbiana. El "bucle microbiano" se refiere a las pérdidas de DOC de todos los organismos y su reintroducción en la cadena alimentaria por bacterias heterótrofas y protistas.

El bacterioplancton también juega un papel importante en otros ciclos biogeoquímicos, tales como el del nitrógeno, fósforo o azufre y en procesos anaeróbicos como la fermentación (Kirchman, 2000).

BACTERIAS HETEROTRÓFICAS MARINAS: El filo Bacteroidetes

En el océano coexisten diferentes grupos de bacterias heterótrofas. Los grupos más abundantes son las Proteobacterias, en particular las clases Alphaproteobacteria (González y Moran, 1997) y Gammaproteobacteria (Giovannoni y Rappé, 2000), las

cuales representan el 30-40% de las células procariotas, y los Bacteroidetes, que representan el 10 - 30% (Cottrell y Kirchman, 2000a; Glöckner *et al*., 1999.). El filo Bacteroidetes es muy diverso y está compuesto principalmente por tres grandes clases ampliamente distribuidas, incluyendo suelos, sedimentos, fuentes hidrotermales, agua dulce y salada y el intestino y la piel de los animales: Bacteroidia, Flavobacteria y Sphingobacteria.

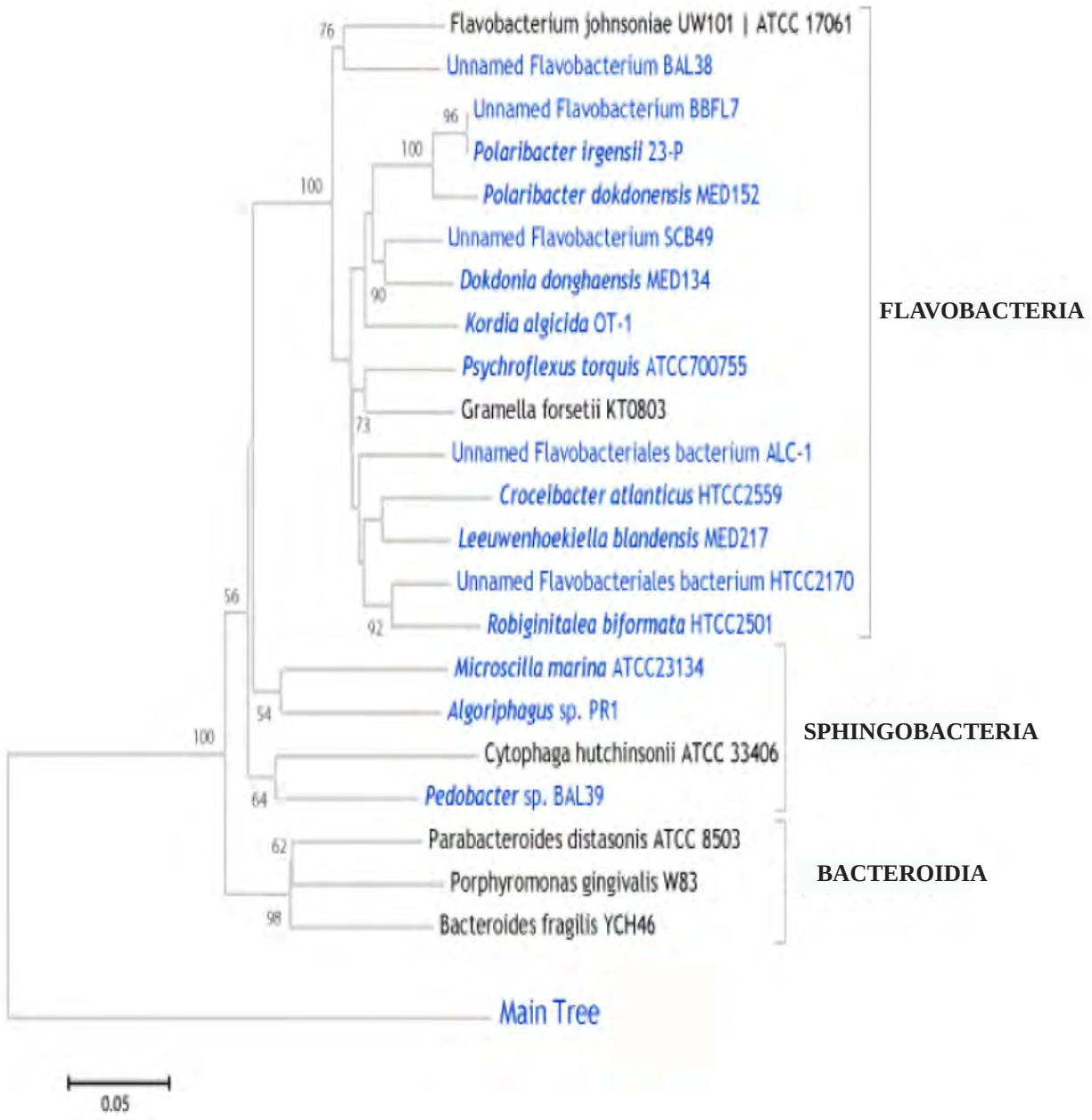


Figura 3. Árbol filogenético de los principales clases dentro del filo de los Bacteroidetes (Página web de la Fundación G. & B. Moore).

La clase Bacteroidia es la mejor estudiada en microbiología general, debido a sus implicaciones para la salud humana (Finegold *et al*., 1983). Ésta incluye el género *Bacteroides* (el más abundante en el intestino de los animales de sangre caliente, incluidos los humanos) y *Porphyromonas*, un grupo de organismos que habitan en la cavidad oral humana. Sin embargo, las clases Flavobacteria y Sphingobacteria, cuyos representantes son en su mayoría marinos y de suelos, sólo han recibido atención en las últimas décadas.

Las Flavobacterias, la clase más abundante del filo Bacteroidetes en el mar (Alonso *et al*., 2007), son degradadores de polímeros y materia orgánica particulada (Pinhassi *et al*., 2004;. Riemann *et al*., 2000). Muchos representantes poseen motilidad por reptación y son importantes habitantes de agregados marinos (DeLong *et al*., 1993;. Kirchman, 2002), pudiendo utilizar las sustancias poliméricas como fuente de carbono y energía.

Durante las proliferaciones de algas las flavobacterias están sobrerepresentadas respecto a otros grupos como respuesta al carbono orgánico presente, lo que denota su predilección por moléculas grandes, y concuerda con su supuesto papel como mineralizadores de la materia orgánica (Cottrell y Kirchman, 2000b). Esta característica contrasta con las de otros grupos importantes de bacterias como las Proteobacterias. Tanto las Alfa- como las Gammaproteobacteria parecen preferir los monómeros a los polímeros y llevar una existencia de vida libre en la columna de agua, sugiriendo una complementariedad en la degradación completa de la materia orgáncia (Cottrell y Kirchman, 2000b).

Uno de los objetivos del presente trabajo fue estudiar la ecología y fisiología de algunos de los representantes de las flavobacterias.

FUENTES PRINCIPALES DE CARBONO PARA EL BACTERIOPLANCTON

Aunque se sabe que el agua de mar contiene materia orgánica en una serie continua de tamaños (Sharp, 1973) interconectados entre ellos (Verdugo *et al*., 2004), esta materia orgánica ha sido pragmáticamente dividida en dos grupos (disuelta y particulada) en base a la filtración a través de filtros GF/F.

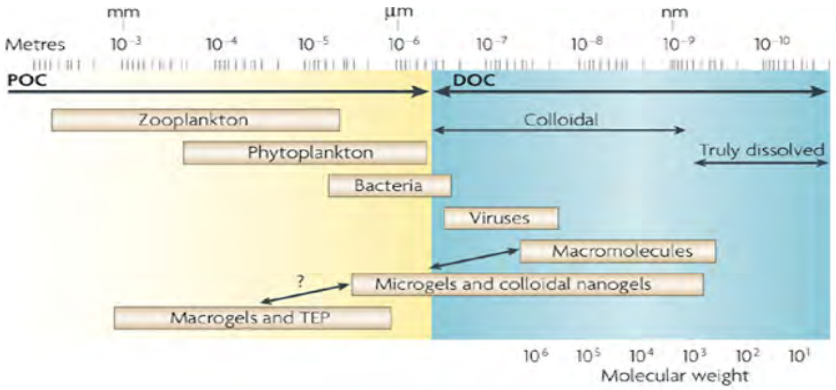


Figura 4. Rango de medidas de la materia orgánica e interacciones microbianas en el océano (Azam and Malfatti 2007).

La principal forma de carbono orgánico (>97%) en el agua de mar está representada por la DOM (Benner, 2002), que se produce principalmente por la exudación de algas (Nagata, 2000) y la lisis celular (infecciones virales) (Weinbauer y Peduzzi, 1995). Sin embargo, las bacterias también producen polisacáridos en la superficie celular que pueden expulsar al medio. Gran parte de esta DOM se compone de moléculas refractarias de carbono que las bacterias no son capaces de asimilar y que permanecen en el océano durante miles de años (Bauer *et al*., 1992; Bauer *et al*., 2002; Williams y Druffel, 1987). Pero también hay una pequeña fracción (<10%) de moléculas disponibles (lábiles) (mono- y polisacáridos, proteínas, péptidos y aminoácidos) (Benner, 2002) que las bacterias pueden utilizar fácilmente. La DOM es liberada principalmente en forma de moléculas de alto peso molecular (HMW-DOM,>1kDa), preferida por las bacterias por su elevada bioreactividad en comparación con el material refractario de bajo peso molecular (LMW-DOM, Amon y Benner, 1994). Sin embargo, las bacterias sólo pueden incorporar pequeñas moléculas en la célula a través de canales transmembrana. Con el fin de asimilar las moléculas grandes, primero tienen que romperlas mediante determinadas enzimas hidrolíticas extracelulares (glucosidasas, peptidasas, lipasas o quitinasas) (Arnosti *et al*., 2005; Hollibaugh y Azam, 1983; Martínez *et al*., 1996).

La materia orgánica que es retenida en filtros de fibra de vidrio de aproximadamente 0.7µm de tamaño de poro constituye la materia orgánica particulada (POM). Ésta es generalmente más biodegradable que la dissuelta (Amon y Benner, 1996; Hedges, 1992). La POM incluye tanto los componentes vivos (plancton) como los no vivos (detritus) (Nagata, 2008). La mayoría de la POM se compone de pequeñas partículas

en forma de geles que flotan en agua de mar y se originan a partir de la agregación de coloides presentes en el océano (que representan el 10-40% de la DOM) (Chin *et al*., 1998;. Kepkay, 1994). Estas macromoléculas tienden a agregarse y formar macro-geles que pueden llegar a alcanzar un tamaño en el que comienzan a sedimentar (Engel *et al*., 2004). Cuando los agregados tienen un tamaño macroscópico (> 500µm) se les conoce como *"nieve marina*" (Suzuki N. y Kato K., 1953). La *"nieve marina"* se refiere a la agregación de partículas grandes y su unión con otros materiales como el fitoplancton o detritus. Se ha demostrado que estos agregados albergan comunidades bacterianas que difieren de las de la columna de agua (Alldredge *et al*., 1986). Las bacterias adheridas a la nieve marina poseen una alta actividad hidrolítica (Smith *et al*., 1992), dejando un rastro de DOM que otras bacterias pueden utilizar.

Las Flavobacterias marinas son especialmente expertas en degradar moléculas de alto peso molecular-DOM (Cottrell y Kirchman, 2000b; Cottrell *et al*., 2005; Kirchman, 2002) y se encuentran con más frecuencia asociadas a la nieve marina en lugar de llevar una vida libre en el agua (Alldredge *et al*., 1986.; DeLong *et al*., 1993;. Johannes *et al*., 1998). Debido a las implicaciones que tiene la colonización de la nieve marina en el ciclo del carbono, los Bacteroidetes marinos tienen un papel importante en la estructuración del ecosistema marino.

En mi tesis analicé varios genomas de Flavobacterias, en busca de posibles enzimas de degradación (glicosidasas y peptidasas) y otros genes que pudieran corroborar su predilección por las macromoléculas y su asociación con agregados. El propósito de este estudio era comprobar si el papel postulado para estas bacterias en los estudios mencionados era coherente con su dotación genética.

FOTOHETEROTROFÍA EN EL OCÉANO: Proteorodopsinas

Entre las bacterias heterótrofas, varios grupos incluyen miembros que poseen mecanismos para obtener energía de la luz solar, un proceso metabólico conocido como fotoheterotrofía. Se conocen principalmente dos sistemas captadores de luz que permiten fotoheterotrofía en el océano: La bacterioclorofila *a* en bacterias fotosintéticas

aerobicas anoxigénicas (AAPs) y las proteorodopsinas (PR), en representantes de las ß-proteobacterias (ej. *Methylophilales* sp.), α-proteobacterias (ej. *Pelagibacter* sp.), ϒ-proteobacterias (ej. *Vibrio* sp.), Flavobacteria (ej. *Dokdonia* sp.) y Planctomycetes (McCarren y DeLong, 2007).

Las proteorodopsinas pertenecen a la familia de las rodopsinas bacterianas, unas proteínas transmembrana que se unen al retinal y funcionan como una bomba de protones impulsada por la luz (Béjà *et al.*, 2000). Son homólogas a las bacteriorodopsinas anteriormente descubiertas en arqueas que habitan los cristalizadores de salinas solares (Oesterhelt y Stoeckenius, 1973). Ambos tipos de rodopsinas actúan como bombas de protones que generan un potencial de membrana quimiosmótico a partir de la energía de la luz, que finalmente puede utilizarse para generar adenosina trifosfato (ATP) (Fig. 5).

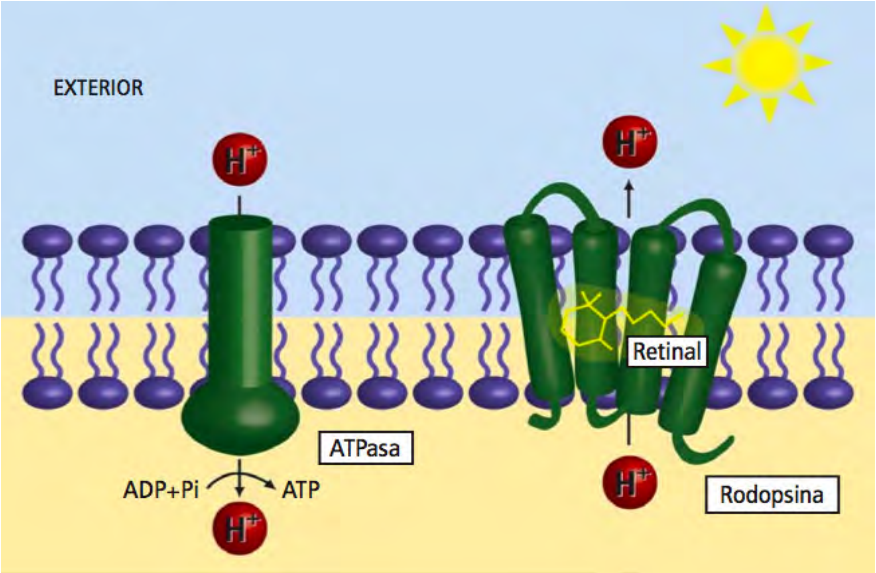


Figura 5. Esquema del sistema previsto para la síntesis de ATP dependiente de proteorodopsina. José M. González (Investigación y Ciencia).

Su descubrimiento (Béjà *et al.*, 2000) fue una verdadera revolución, ya que demostraba la existencia nuevos metabolismos fototróficos de los que se conocían hasta el momento basados en moléculas del tipo clorofila en el mar.

Estudios posteriores detectaron una amplia gama de variantes de genes de la PR y su distribución entre numerosos grupos bacterianos (Campbell *et al.*, 2008; de la Torre *et al.*, 2003) y arqueas (Frigaard *et al.*, 2006.). La iniciativa de secuenciación

genómica de microorganismos marinos llevada a cabo por la Fundación Gordon y Betty Moore (www.moore.org) permitió la secuenciación de más de 170 genomas de bacterias marinas. El análisis de estos genomas reveló la presencia de proteorodopsinas en varios grupos bacterianos tales como las Gammaproteobacteria (ej., *Vibrio campellii* AND4, *Vibrio angustum* S14), Alphaprotebacteria (ej., *Photobacterium* sp. SKA34, *Pelagibacter ubique* HTCC1062), Betaproteobacteria (ej., Methylophilales bacterium HTCC2181) y Bacteroidetes (ej., *Dokdonia* sp. MED134, *Polaribacter* sp. MED152, *Polaribacter irgensii* 23-P, Flavobacterium bacterium BAL38).

A pesar de la supuesta ventaja adaptativa de poseer el gen de la PR, no existe evidencia clara sobre el papel metabólico de las proteorodopsinas, principalmente debido al hecho de que sólo se han cultivado unas pocas bacterias con PR con éxito (Fuhrman *et al*., 2008; Riedel *et al*., 2010).

Sólo *Dokdonia* sp. MED134, aislada en el Observatorio Microbiano de la Bahía de Blanes, ha mostrado mayores tasas de crecimiento en luz comparadas con la oscuridad en situación de limitación de materia orgánica (Gómez-Consarnau *et al.*, 2007; Kimura *et al*., 2011). Esto significa que MED134 podría obtener energía de la luz en condiciones oligotróficas y sobrevivir hasta colonizar nuevas partículas a las que adherirse. Queda por demostrar si esto se produce también en las comunidades naturales de aguas oligotróficas.

En el presente trabajo he investigado si la existencia de PR supone un verdadero beneficio para las bacterias que la poseen en condiciones de baja concentración de materia orgánica.

GENÓMICA FRENTE A CULTIVOS

Desde que la llamada "revolución genómica" comenzó hace unos 15 años, la genómica y la metagenómica se han convertido en herramientas poderosas para explorar el mundo microbiano (Xu, 2006). Estas herramientas tienen el potencial de revolucionar nuestra comprensión sobre el papel de los microorganismos en el océano (Pedrós-Alió, 2006).

El estudio de todos los genes codificados en el material genético que posee un organismo fue posible gracias a los avances tecnológicos en el campo de la biología molecular (Moran *et al*., 2004a). Al principio, la capacidad de secuenciación permitió determinar la diversidad y distribución de algunos genes seleccionados. Por ejemplo, la amplificación y secuenciación del rRNA 16S o del gen *nif*H proporcionó información sobre las relaciones filogenéticas entre las especies (Woese *et al*., 1975) y sobre el metabolismo de fijación de nitrógeno de las bacterias (Zehr y Capone, 1996), respectivamente.

Estos estudios moleculares difieren de la genómica en el hecho de que ésta última permite analizar más genes a la vez y posibilita encontrar nuevos genes sin tener que conocerlos previamente. Se estima que se han aislado en cultivo puro menos del 1% de todas las especies bacterianas (Amann *et al*., 1990). Aún así, los microorganismos cultivados posibilitan el estudio de sus familiares no cultivados, convirtiéndose en organismos modelo para comprobar hipótesis acerca de su fisiología. Algunos ejemplos

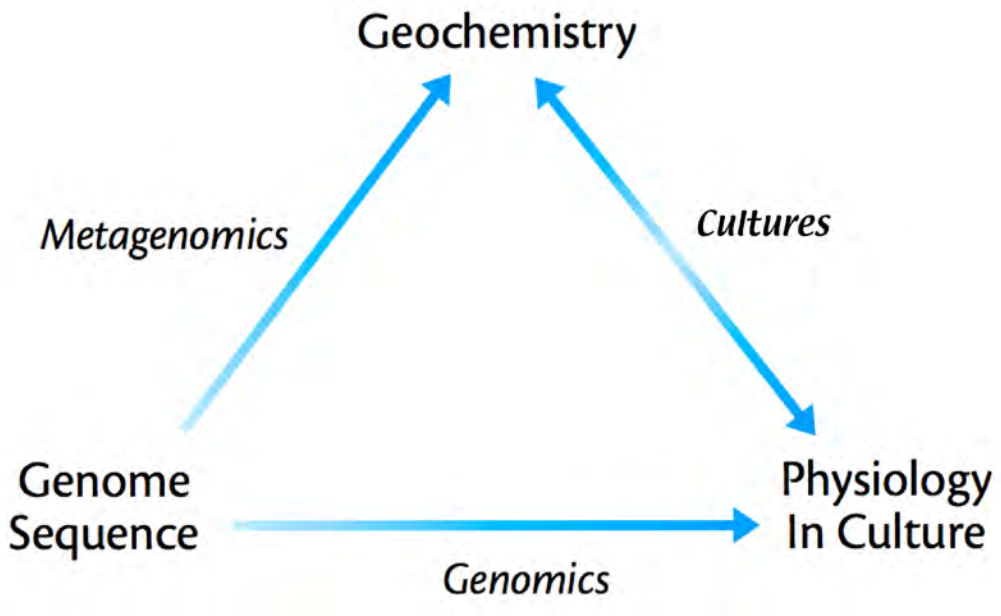


Figura 6. Relación entre genómica, metagenómica y cultivos. (adaptado de Giovannoni *et al*., 2007).

de microorganismos marinos modelo son: *Prochlorococcus marinus* (Rocap *et al*., 2003), *Pelagibacter ubique* (Giovannoni *et al*., 2005), *Silicibacter pomeroyi* (Moran *et al*., 2004b) o *Rhodopirellula baltica* (Glöckner *et al*., 2003). La genómica de microorganismos cultivados está ayudando a clarificar quién está haciendo qué y cómo. Sin embargo, la genómica tiene que hacer frente al hecho de que no todos los microorganismos han sido aislados en cultivo puro.

La metagenómica es distinta, pero complementaria, a la genómica de aislados. También conocida como genómica de comunidades o ambiental, la metagenómica es la secuenciación y el análisis de DNA de todos los microorganismos recuperados de un ambiente evitando la necesidad de aislarlos y cultivarlos individualmente (Chen y Pachter, 2005). El conjunto resultante de genes descifrado a partir de microorganismos no cultivables tiene un tremendo potencial en el desarrollo de biocatalizadores novedosos para aplicaciones industriales y médicas. Dos ejemplos de metagenómica son el estudio del Mar de los Sargazos (Venter *et al*., 2004) donde se generaron un total de 1.045×10^9 pares de bases y el Muestreo Global del Océano (GOS, Rusch *et al*., 2007) que obtuvo un amplio conjunto de datos consistente en aproximadamente 6.3×10^9 pares de bases. Otro ejemplo es el análisis de bibliotecas de clones con fragmentos grandes de DNA en cada clon. Fue ésta aproximación la que permitió descubrir los genes de la PR (Béjà *et al*., 2000). Otros estudios revelaron la presencia del gen de la amonio monooxigenasa (para la oxidación de amonio, produciendo nitrito) en arqueas, provocando la reevaluación de la idea preconcebida de que la nitrificación en el océano era una función de carácter exclusivamente bacteriana (Nicol y Schleper, 2006).

A pesar de los notables logros que ha proporcionado la metagenómica en la predicción de las actividades metabólicas microbianas a partir de secuencias del genoma, el cultivo sigue siendo una técnica fundamental y útil que permite el estudio de las características fisiológicas de los microorganismos. El estudio de los microorganismos en cultivo puro es la única manera de probar las hipótesis generadas por los análisis *in silico* de secuencias o estudios ecológicos (Giovannoni *et al*., 2007). La figura 6 muestra las relaciones entre las diferentes disciplinas mencionadas en este párrafo: La genómica aumenta la cantidad de información obtenida de un genoma, permitiendo una planificación óptima de los estudios fisiológicos que se pueden realizar en el laboratorio. Estos estudios fisiológicos, a su vez, aumentan la calidad en la predicción de las actividades geoquímicas de las células. La metagenómica elude la necesidad de cultivar, pero enton-

ces, las hipótesis surgidas de los genomas son más difíciles de probar. La ventaja es que la metagenómica revela lo que existe en la naturaleza sin necesidad de saberlo de antemano.

La mayor parte de la presente tesis se realizó usando aislados marinos como una forma de investigar y comparar el potencial de la genómica en los Bacteroidetes. El cultivo fue un intento de corroborar la información obtenida de los genomas.

LA TRANSFERENCIA HORIZONTAL DE GENES ORIGINA ISLAS GENÓMICAS

Uno de los mecanismos mediante los cuales las bacterias evolucionan es la adquisición de nuevos genes, los cuales pueden ser beneficiosos en ciertas condiciones ambientales, a través de la transferencia horizontal de genes (HGT) (Dobrindt *et al*., 2004). Este mecanismo permite a las poblaciones bacterianas diversificarse rápidamente y adaptarse a nuevos nichos ecológicos (Boucher *et al*., 2003). Muchos de los genes adquiridos por HGT forman grupos de genes conocidos como islas genómicas (IGs). Las IGs son fragmentos de DNA flanqueados por elementos (como transposasas o integrasas) que permiten la movilización del material genético. Las IGs contienen elementos móviles del DNA (MGE: mobile genetic elements) tales como integrones, transposones conjugativos o profagos. Estas islas se caracterizan por tener un tamaño relativamente grande (Hacker y Kaper, 2000), estar asociadas frecuentemente con tRNA (Williams, 2002) y presentar un porcentaje diferente en el contenido de GC en comparación con el resto del genoma (Navarre *et al*., 2006). Cuando se descubrieron por primera vez, estos elementos fueron asociados a patogenicidad (Hacker *et al*., 1990), pero en la actualidad se acepta que las islas genómicas representan una variedad mucho más amplia de funciones que la de la patogenicidad (Dobrindt *et al*., 2004). Las IGs albergan una gran diversidad de genes. Sin embargo, algunas características y patrones generales pueden ser reconocidos mediante la comparación con diferentes genomas bacterianos (Vernikos y Parkhill, 2008). Algunos estudios que describen IGs en genomas microbianos ambientales sugieren que la adquisición de nuevas características metabólicas, mecanismos de defensa, transporte de determinadas sustancias o posibilidad de degradar metales son rasgos comunes

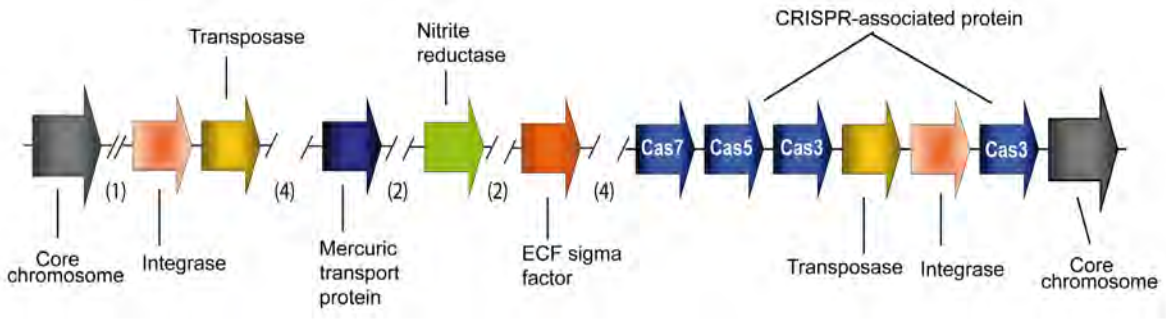


Figure 7. Esquema de una isla genómica modelo en bacterias. Los números entre paréntesis indican las proteínas hipotéticas que no se han considerado para la figura.

que caracterizan a las IGs.

Con el fin de estudiar estas IGs, se han desarrollado programas de predicción de islas genómicas. Existen métodos basados en la secuencia de núcleotidos que dependen de la variación natural que existe entre la composición de la secuencia del genoma, por ejemplo, las regiones donde el contenido de G+C cambia. Sin embargo estos métodos pueden tener dificultad para la detección de las islas debido a la variación normal en la composición de la secuencia que puede ocurrir dentro de los genomas bacterianos. Los nuevos enfoques como los métodos de genómica comparativa se basan en tratar de identificar las regiones transferidas horizontalmente utilizando la información de varias especies relacionadas (Langille *et al.*, 2010).

En la presente tesis, determiné las islas genómicas de algunos representantes de los cuatro grupos principales de bacterias en el océano y los analicé en busca de patrones y genes dentro de estas islas genómicas que pudieran proporcionar mejoras adaptativas al ambiente.

OBJETIVOS DE LA TESIS

El objetivo principal de la presente tesis es el de ampliar el conocimiento que se tenía de los Bacteroidetes marinos (o Flavobacterias), un grupo bacteriano muy abundante e importante en el océano pero quizás uno de los menos estudiados hasta el momento. Pretendemos profundizar más en el estudio de su genética (analizando su genoma y investigando sus rutas metabólicas y genes de interés ecológico), fisiología (realizando experimentos para ver su respuesta frente a determinados factores) y de entender sus estrategias de supervivencia en el océano (pudiéndolas comparar con otros grupos bacterianos importantes). Con el fin de llegar a responder a todas nuestras preguntas, hemos llevado a cabo análisis y experimentos tanto *in silico* como *in vivo*, combinando técnicas bioinformáticas con experimentos de crecimiento. Esta tesis consta de cinco trabajos diferentes organizados en cinco capítulos que tratan de responder preguntas más específicas enumeradas a continuación:

¿Cuál es la composición genómica de *Polaribacter* sp. MED152, un representante de los Bacteroidetes marinos? ¿Qué nos dice el análisis de su genoma acerca de sus estrategias de vida?

En el **Capítulo 1** titulado *Análisis del genoma de Polaribacter sp. MED152, una bacteria marina que contine la protorodopsina,* se analizó el genoma completo de esta flavobacteria aislada en el Observatorio Microbiano de la Bahía de Blanes (BBMO) y que consideramos de interesante estudio por el hecho de poseer el gen de la proteorodopsina. El análisis de genomas completos de cianobacterias y proteobacterias nos habían llevado a comprender sus estrategias de vida y metabolismos. Sin embargo, estudios similares en Bacteroidetes todavía no se habían llevado a cabo.

Así, el genoma de MED152 podría proporcionarnos nuevos conocimientos sobre las capacidades fisiológicas de bacterias con proteorodopsina, convirtiéndose en un genoma modelo para el estudio de procesos celulares y moleculares en bacterias que expresan este gen, su adaptación al medio marino y su papel en el ciclo del carbono.

¿Hay algún patrón entre los Bacteroidetes marinos? Podrían encontrarse características similares a las de *Polaribacter* en otros representantes?

Posteriormente, en el **Capítulo 2** titulado *Genómica comparativa de Bacteroidetes marinos con y sin proteorodopsina*, quisimos saber si las características halladas en el genoma de *Polaribacter* sp. MED152 y en otra flavobacteria cuyo genoma también había sido completamente descrito anteriormente, *Gramella forsetii* KT0803, aislada en el Mar del Norte y que no contiene la proteorodopsina, eran comunes entre ellos y a otros representantes de las flavobacterias y diferentes a otros grupos de bacterias marinas. Para ello seleccionamos otras dos flavobacterias cuyos genomas también habían sido secuenciados, *Dokdonia* sp. MED134 (con proteorodopsina) y *Leewenhoekiella blandensis* MED217 (sin proteorodopsina). Analizamos estos genomas con el objetivo principal de ver qué rasgos genéticos identificados en MED152 estaban restringidos a bacterias que poseen la proteorodopsina, cuales eran comunes a Bacteroidetes marinos y cuales estaban presentes en todos los Bacteroidetes (incluyendo los simbiontes del tracto intestinal de mamíferos).

¿Proporcionan las islas genómicas presentes en algunos grupos de bacterias ventajas para su capacidad ecológica adaptativa? Hay variación en el número y naturaleza de estas islas entre diferentes grupos de bacterias?

Seguidamente, en el **Capítulo 3** que lleva por título *Patrones y arquitectura de la islas genómicas de bacterias marinas*, quisimos ir más allá y decidimos hacer un análisis exhaustivo de las islas genómicas (fragmentos de DNA permiten la movilización del material genético) de los Bacteroidetes marinos, comparándolas con las de otros grupos de bacterias tales como las cianobacterias, las alfa- y gamma-proteobacterias y los Bacteroidetes no marinos. Usando la base de datos IslandViewer y los programas de detección de islas que tiene integrados en su interfaz, llevamos a cabo el análisis de las islas genómicas de 70 genomas y construimos una base de datos con un total de 8152 genes presentes en las 436 islas genómicas detectadas. En este estudio exploramos la distribución, patrones y función de los genes presentes en las islas, lo cual nos ha proporcionado una nueva visión de la relevancia ecológica de estos elementos móviles en los genomas procariotas marinos.

¿Cuál es la función real de los genes predichos (concretamente la del gen de la PR) en el medio ambiente natural? ¿Pueden las técnicas tradicionales de cultivos puros ayudarnos a descubrir esto?

En el **Capítulo 4** que lleva por título *El medio de cultivo influencia la respuesta a la luz de Flavobacterias marinas que contienen proteoprodopsina,* quisimos centrarnos en la posible ventaja adaptativa que proporcionaría el hecho de poseer el gen de la proteorodopsina. Anteriormente, en el **Capítulo 1**, ya habíamos formulado hipótesis de los posibles beneficios de poseer este gen pero hasta el momento muy pocos estudios se habían realizado para comprobarlo experimentalmente. Los objetivos de este capítulo eran determinar la influencia de la luz en el crecimiento y en la incorporación de bicarbonato en MED152 y en MED134. Para ello diseñamos una serie de medios pobres en materia orgánica, con diferentes concentraciones de carbono orgánico disuelto para ver si se producía estimulación del crecimiento y de la asimilación de carbono inorgánico en presencia de luz con respecto a la oscuridad.

¿Cómo es posible determinar la capacidad metabólica del bacterioplancton en el medio natural?

Finalmente, en el **Capítulo 5** llamado *Cambios estacionales en los patrones de utilización de substratos por el bacterioplancton en el Golfo de Amundsen (Ártico Occidental)*, quisimos estudiar el patrón de utilización de substratos de la comunidad microbiana el Golfo de Amundsen (Mar de Beaufort) de invierno (febrero) a verano (julio). El área de estudio es una zona muy dinámica y con unas condiciones de hielo muy heterogéneas, conocida como polynya. Para ello diseñamos e inoculamos microplacas BIOLOG MT2 con diferentes fuentes de carbono (polímeros, carbohidratos, aminacidos, y otros substratos) las cuales nos proporcionaron información sobre los substratos más utilizados por el bacterioplankton heterotrófico. Estas microplacas constan de 96 pocillos vacios que contienen violeta de tetrazolio, que se reduce de forma irreversible a formazán como consecuencia de la actividad metabólica bacteriana dando un color violeta en el pocillo que facilita la lectura visual de los resultados.

Nuestros resultados los pudimos comparar con los obtenidos en un estudio anterior que se realizó en la misma zona pero en un área con una hidrografía completamente diferente, mostrando una capa de hielo estable a lo largo de todo el año.

METODOLOGÍA

En el presente trabajo, se han utilizado herramientas *in silico* para llegar a la mayoría de los objetivos anteriormente citados (**Capítulos 1-3**). También se han realizado experimentos *in vivo* con cultivos (**Capítulo 4**) para comprobar algunas de las hipótesis que surgieron a partir del análisis genómico previamente realizado. Por último, se utilizó una técnica de huella digital metabólica (microplacas Biolog) para caracterizar la comunidad microbiana *in situ* (**Capítulo 5**). Aunque la metodología particular de cada capítulo está detallada en los mismos, en esta sección se presenta una visión global de las diferentes técnicas. El siguiente diagrama muestra los diferentes criterios utilizados en el estudio de bacterias marinas y sus interconexiones.

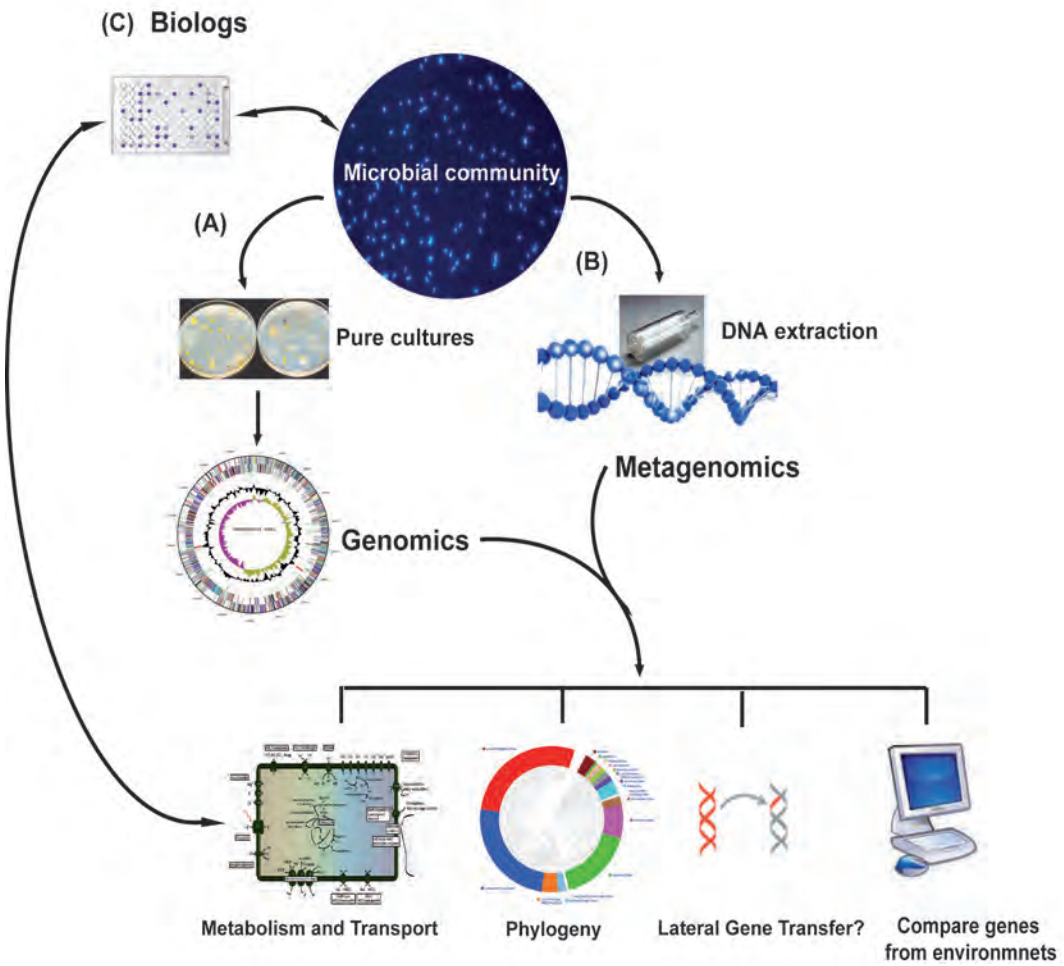


Figura 8. Diferentes enfoques para el estudio de la comunidad microbiana utilizada en esta tesis. (A) Un organismo puede ser aislado en cultivo puro. La secuenciación de su genoma permite estudiar el DNA, mRNA y proteínas para obtener información acerca de: i) filogenia, ii) metabolismo y transporte, iii) existencia de transferencia lateral de genes, y

iv) similitudes con organismos de otros ambientes. **(B)** Una información similar puede ser obtenenida estudiando toda la comunidad microbiana a través de metagenómica, evitando la necesidad de cultivar los organismos (en la presente tesis hemos elegido la vía **A**). **(C)** Las microplacas BIOLOG permiten estudiar algunas funciones metabólicas en muestras tomadas directamente de la comunidad. Algunos de los rasgos que se encuentran por medio de la genómica y metagenómica también se pueden probar usando estas placas.

Aproximanciones "*in silico*": HERRAMIENTAS BIOINFORMÁTICAS

Anotación genómica y análisis

La anotación genómica es el proceso de asignación de la información biológica a las secuencias de genes y sus productos proteicos (Stein, 2001). Durante los últimos años, se han secuenciado un gran número de genomas (Overbeek *et al*., 2004) y la anotación manual de estos genomas es una tarea que consume mucho tiempo. Para resolver este cuello de botella, se han desarrollado cadenas de procesos (pipelines) para automatizar la anotación; estas cadenas realizan búsquedas de similitud seguidas de una evaluación automática de los resultados y la generación de una anotación funcional. La anotación automática acelera el proceso. Sin embargo, una anotación de alta calidad requiere un refinado manual de las anotaciones. Hay varios portales que ofrecen herramientas y servicios para la anotación automática y su revisión manual. En la presente tesis (**Capítulo 1**), se utilizó el portal GenDB del Centro de Biotecnología (CeBiTec) de la Universidad de Bielefeld (Alemania).

El gráfico siguiente muestra el flujo de datos en la anotación genómica. El proceso comienza con la secuencia terminada del genoma y la predicción de genes y sus funciones. Hay una retroalimentación entre la identificación de genes y la corrección de errores de secuenciación. Después de la predicción automática de genes, la búsqueda de similitud entre genes se lleva a cabo mediante herramientas de alineamiento de secuencias (**Basic Local Alignment Search Tool**, BLAST) (Altschul *et al*., 1990), que identifica los casos con alta similitud con otras proteínas en bases de datos (normalmente NCBI NR), de modo que se puede predecir la función. Se puede realizar una anotación más detallada utilizando otras herramientas en combinación con la búsqueda de bases de datos especializadas como KEEG (Kyoto Encyclopedia of Genes and Genomes, Kanehisa and Goto, 2000) o Pfam (familia de parálogos, Finn *et al*., 2008).

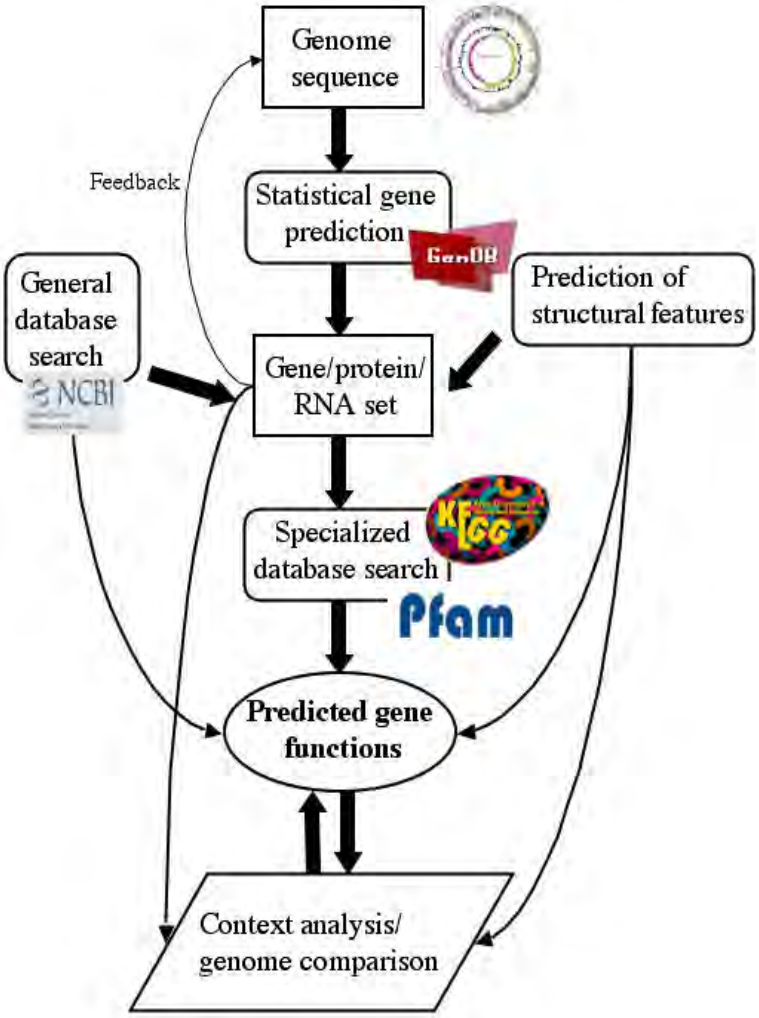


Figura 9. Diagrama de flujo generalizado del proceso de anotación genómica.

Comparación genómica

Si bien el análisis de un genoma único ofrece enormes conocimientos biológicos sobre cualquier organismo, el análisis comparativo de múltiples genomas proporciona mucha más información sobre la fisiología y la evolución de las especies microbianas y amplía la capacidad para asignar una mejor función putativa a los genes predichos. El análisis comparativo ofrece una nueva perspectiva en las relaciones entre los genes homólogos (Koonin *et al.*, 2005). Comparando secuencias de genes entre los genomas y dentro de cada genoma, es posible reconstruir la historia evolutiva de cada gen. Esto, a

su vez, permite una mejor comprensión de las adaptaciones específicas de los genomas (Koonin *et al*., 2005).

La genómica comparativa implica el uso de programas que puedan alinear numerosos genomas y buscar regiones semejantes entre ellos. Algunas de estas herramientas de similitud de secuencias son accesibles al público a través de Internet, el más utilizado es el ya mencionado BLAST. En el **Capítulo 2** se utilizó el criterio denominado **Reciprocal Best Match** (RBM) para encontrar ortólogos (genes homólogos con la misma función relacionados a través de la especiación) y parálogos (genes homólogos relacionados a través de la duplicación y que evolucionan originando nuevas funciones) entre y dentro de diferentes representantes de Bacteroidetes marinos. Esto no es más que un BLAST en dos sentidos, el gen *a* contra el gen *b* y viceversa, y si la similitud es buena (según los parámetros introducidos) en ambas comparaciones, entonces se consideran genes homólogos.

Con el fin de explorar y visualizar fácilmente los resultados, se utilizó el programa JCoast, www.jcoast.net (Richter *et al*., 2008), que permite encontrar rápidamente un gen de interés y compararlo con sus ortólogos en otros genomas.

Predicción de Islas Genómicas

Es posible detectar islas genómicas (IG) gracias a los diversos programas que recientemente se han desarrollado. En el **Capítulo 3** elegimos IslandViewer, una interfaz integrada para la identificación y la visualización de las islas genómicas (Langille y Brinkman, 2009; Langille *et al*., 2010). IslandViewer reúne tres métodos diferentes: SIGI-HMM (mide el uso de codones), IslandPath-DIMOB (detecta en la irregularidades en composición de la secuencia o la presencia de elementos móviles), e IslandPick (utiliza la genómica comparativa). Este sitio en la red contiene los resultados pre-calculados para un gran número de genomas secuenciados y permite al usuario enviar genomas para ser ananlizados.

Experimentos "*in vivo*"

Cultivos

Como complemento a la genómica, en el **Capítulo 4**, se realizó un considerable esfuerzo para desarrollar un medio de cultivo definido donde los Bacteroidetes pudiesen crecer con el fin de examinar sus adaptaciones fisiológicas.

Microplacas BIOLOG MT

Estas placas se crearon para análisis de comunidades microbianas y estudios de ecología microbiana (Garland y Mills, 1991; Insam, 1997), permitiendo determinar las características fisiológicas de la comunidad y distinguir cambios espaciales y temporales. Las placas tienen 96 pocillos y contienen una fuente de carbono y l violeta de tetrazolio como un indicador redox de utilización de la fuente de carbono. Cuando las bacterias oxidan el compuesto, el indicador cambia a color púrpura. Los usamos en el **Capítulo 5** en un intento de detectar las actividades de degradación de polímeros de una comunidad microbiana.

RESUMEN Y RESULTADOS

CAPÍTULO 1

Análisis del genoma de *Polaribacter* sp. MED152, una bacteria marina que contine la proteorodopsina

El análisis de genomas de cianobacterias y proteobacterias marinas ha proporcionado un profundo conocimiento acerca de las estrategias de vida de estos organismos, su diferenciación en ecotipos y su metabolismo. Sin embargo, todavía falta un análisis comparable para los Bacteroidetes, el tercer mayor grupo de microorganismos del plancton marino. En este capítulo presentamos el análisis completo del genoma de *Polaribacter* sp. MED152.

Encontarmos que, por un lado, MED152 contiene un número importante de genes para la adhesión a superficies o partículas, motilidad por reptación y degradación de polímeros. Esto concuerda con la estrategia de vida que en la actualidad se le da a los Bacteroidetes marinos. Por otro lado, también observamos que este microorganismo contiene el gen de la proteorodopsina, junto con un notable conjunto de genes para detectar y responder a la luz, que podría proporcionar una ventaja para su supervivencia en la superfície iluminada de océanos pobres en nutrientes, en la búsqueda de nuevas partículas que colonizar.

Por otra parte, vimos aumento en la fijación de CO_2 en presencia de luz, cosa que indicaba que el metabolismo central limitado podría complementarse con la fijación anaplerótica de carbono inorgánico, mediante una combinación única de transportadores de membrana y carboxilasas.

Esto sugiere una estrategia de doble vida que, si se confirma experimentalmente, sería notablemente diferente de lo que se conoce en los otros dos grupos principales de bacterias (cianobacterias autótrofas y proteobacterias heterótrofas) en la superficie del océano. El genoma de *Polaribacter* ha proporcionado nuevos conocimientos sobre las capacidades fisiológicas de las bacterias que contienen proteorodopsina. Este genoma

servirá como modelo para estudiar los procesos celulares y moleculares de las bacterias que expresan proteorodopsina, su adaptación al medio ambiente oceánico, y su papel en el reciclaje de carbono.

CAPÍTULO 2

Genómica comparativa de Bacteroidetes marinos con y sin proteorodopsina

El análisis comparativo de genomas de Bacteroidetes marinos nos ha revelado una estrategia de vida diferente a las de otros filos importantes del bacterioplancton marino, como las cianobacterias y las proteobacterias, indicando que estas bacterias tienen muchas adaptaciones para crecer adheridas a partículas y degradar polímeros, incluyendo un gran número de peptidasas, glicosil hidrolasas, glisosil transferasas, proteínas de adhesión y genes de motilidad por reptación. Varios de los genes se encuentran en estrecha asociación con receptores dependientes de TonB y transductores de señales, lo que sugiere una regulación integrada para la adherencia y degradación de los polímeros.

Entre los genes compartidos por los cuatro genomas de Bacteroidetes se encuentran genes para la transducción de señales, transporte, movilidad y un gran porcentaje de proteínas de membranas y de secreción. Además, los Bacteroidetes marinos que contienen proteorodopsina parecen cambiar su metabolismo para poder sobrevivir cuando flotan libremente en la superficie del océano pobre en nutrientes. Los genes potencialmente importantes en esta alternancia de estilo de vida incluyen muchos genes implicados en la fijación anaplerótica de CO_2.

Este estudio también demuestra que los Bacteroidetes marinos estan especializados en el uso de materia orgánica polimerica, especialmente proteinas.

CAPÍTULO 3

Patrones y Arquitectura de Islas Genómicas en Bacterias Marinas

Las islas genómicas (IGs) tienen un papel importante en la modulación de la estructura y tamaño del genoma bacteriano, exhibiendo un diverso conjunto de genes lateralmente transferidos. Sin embargo, las islas genómicas de los genomas bacterianos marinas no se han explorado de manera sistemática para descubrir posibles patrones y analizar su importancia ecológica putativa.

En este estudio realizamos el análisis exhaustivo de las islas genómicas presentes en 70 genomas de bacterias marinas (incluyendo Cianobacterias, Alpha- y Gamma-proteobacterias y Bacteroidetes) seleccionados para explorar su distribución, patrones y el contenido funcional de genes en estas regiones genómicas. A través de varios métodos para la detección y visualizción de islas genómicas integrados en la base de datos llamada IslandViewer (http://wwwpathogenomics.sfu.ca), detectamos 438 islas genómicas que contenían un total de 8152 genes. El número de islas genómicas por genoma demostró estar firme y positivamente correlacionado con el tamaño total de islas genómicas. La fracción del genoma bacteriano representado por islas genómicas variaba significativamente entre grupos taxonómicos e incluso dentro de las mismas especies bacterianas. En el 50% de los genomas analizados las islas genómicas representaron aproximadamente el 3% de la longitud del genoma, con un máximo del 12%.

Describimos islas genómicas ricas en elemenos genéticos móbiles y originadas por transferencia horizontal de genes y descubrimos otro tipo de islas originadas por recombinación homóloga en varios géneros de Bacteroidetes marinos y en cepas de *Shewanella*, entre otras. Este segundo tipo de IG sorprende por poseer genes "*housekeeping*" como la subunidad ß de la polimerasa de RNA dirigida por DNA, factores sigma de regulación, el factor de elongación Tu y genes de proteínas ribosomales típicamente asociados con el resto del genoma.

La anotación funcional de todos los genes pertenecientes a islas genómicas reveló un rango muy diverso de genes con relevancia biológica y que además, cada grupo tenía

un conjunto diferente de genes en sus IGs, indicando diferentes estrategias de vida.

Así, nuestros resultados indican que tanto la transferencia horizontal de genes mediada por fagos y otros elementos genéticos móviles desempeñan un papel importante en la movilidad de grupos de genes entre taxones y dentro de genomas estrechamente relacionados, modulando el *pool* flexible del genoma. Nuestros hallazgos sugieren que las islas genómicas pueden aumentar la aptitud de bacterias bajo condiciones cambiantes del medio ambiente mediante la adquisición de nuevos genes y/o modulando la transcripción y/o transducción de genes.

CAPÍTULO 4

El medio de cultivo influencia la respuesta a la luz de las flavobacterias que poseen proteorodopsina

La proteorodopsina (PR) es una proteína fotoactiva presente en el bacterioplancton marino, incluyendo bacterias, arqueas y eucariotas. Se compone de una proteína transmembrana ligada a una molécula de retinal y funciona como una bomba de protones impulsada por la luz. A pesar de ser muy abundante en el océano, su función fisiológica aún no está bien comprendida, ya que se ha sugerido que podría desempeñar un papel diferente en cada una de las cepas estudiadas que contienen este gen. Estudios recientes han demostrado que la luz aumenta el rendimiento celular y la fijación de carbono en diferentes bacterias (*Dokdonia* sp. MED134 y *Polaribacter* sp. MED152, respectivamente) que albergan el gen que codifica para la PR, sugiriendo que podría haber diferentes estrategias para el uso de la PR para cada microorganismo.

En este estudio hicimos experimentos de crecimiento en medios pobres para ver una vez más el efecto de la luz pero esta vez en un medio pobre definido (utilizando extracto de levadura y glucosa como fuente de carbono). Demostramos por primera vez que la luz estimulaba el rendimiento celular en condiciones de crecimiento con baja concentración de carbono también en *Polaribacter* sp. MED152. Confirmamos asimismo que esta mejora no sólo dependía de la concentración, sino también de la naturaleza de la materia orgánica lábil disponible, ya que no se producía cuando, bajo la mismas condiciones de incubación, la fuente de carbono cambiaba.

CAPÍTULO 5

Cambios estacionales en los patrones de utilización de sustratos por el bacterioplancton en el Golfo de Amundsen (Ártico Occidental)

Este trabajo fue realizado durante el estudio llamado Circumpolar Flaw Lead (CFL) que tuvo lugar en el Golfo de Amundsen (Ártico Occidental) con motivo de la celebración del Año Polar Internacional. Ésta es un área muy dinámica con unas condiciones de hielo heterogéneas.

Con este estudio quería mos rastrear la diversidad metabólica del bacterioplancton heterotrófico, investigando su respuesta frente a la adición de diferentes fuentes de carbono. Para ello utilizamos, microplacas BIOLOG MT2® que fueron inoculadas con agua de mar en 17 estaciones. Las muestras fueron tomadas entre febrero y julio de 2008 en superfície (0-12 m) y en profundidad (20-70 m).

En invierno el número de sustratos utilizados fue muy bajo. En primavera, cuando los niveles de clorofila *a* aumentaron, también aumentó el número de sustratos utilizados. Esto fue seguido por un período de un mes sin actividad visible en abril y mayo. Finalmente, la actividad se recuperó hacia verano. Los hidratos de carbono fueron los sustratos más utilizados, especialmente la glucosa y la celobiosa. Los aminoácidos se utilizaron de vez en cuando, tanto en aguas de superficie como en profundidades intermedias. Estos resultados se compararon con los de un estudio previo en la Bahía de Franklin, un área relativamente cerrrada, pero con un régimen de hielo diferente, mostrando una capa estable de la tierra rápida de hielo durante todo el período de estudio.

Las diferencias encontradas en las capacidades metabólicas del plancton marino entre los dos estudios fueron probablemente una consecuencia de la hidrografía diferente entre las dos zonas.

SÍNTESIS DE RESULTADOS Y DISCUSIÓN GENERAL

Los Bacteroidetes marinos desde una perspectiva genética y fisiológica

El propósito de la presente tesis es el de aumentar la comprensión acerca de la genómica y la ecofisiología de los Bacteroidetes marinos usando aproximaciones genómicas y cultivos bacterianos. Los Bacteroidetes son uno de los tres grupos más abundantes de bacterias en el océano (Alonso *et al*., 2007, Alonso-Sáez and Gasol 2007). En los últimos años, el número de genomas bacterianos completos ha ido aumentando en bases de datos tales como GenBank. Actualmente, el número de genomas completos de Bacteroidetes representa la mayor proporción (8,9%) después de las Proteobacterias (29,4% gamma- 17,8% y alphaproteobacteria 11,6%) y Firmicutes (22,9%)(Fig. 10A). Además, la Fundación Gordon y Betty Moore ha secuenciado 16 Bacteroidetes marinos de un total de 182 genomas bacterianos marinos (Fig. 10B).

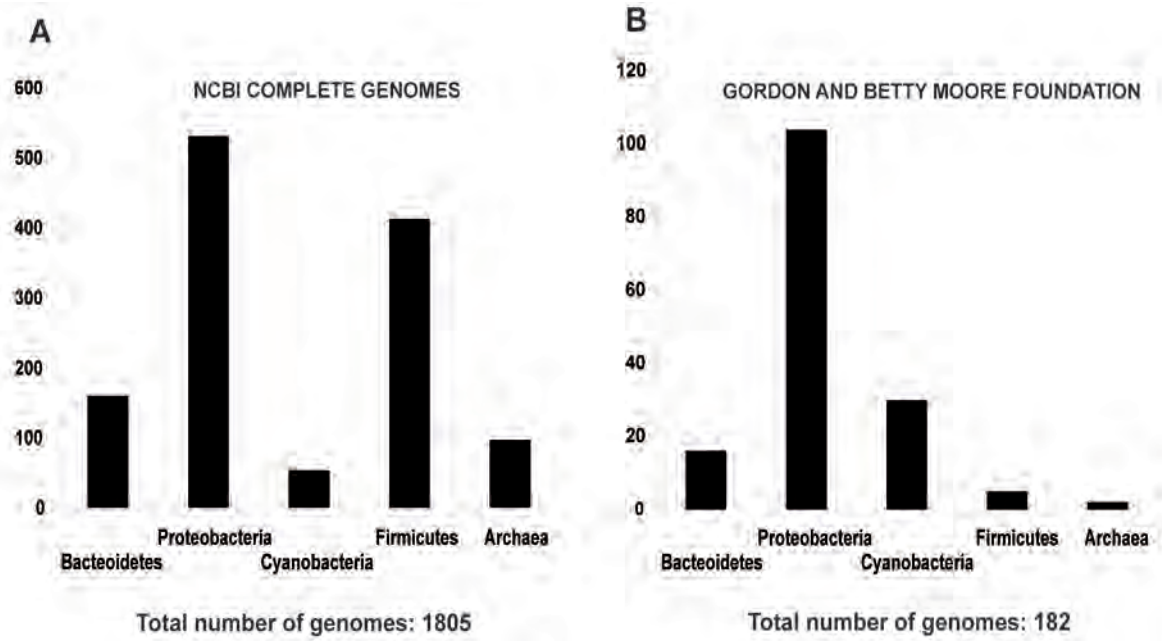


Figura 10. Número de genomas completos de los grupos marinos más abundantes en NCBI (GenBank) y la Fundación G. & B. Moore (actualizada 07052011).

Spanish Summary

En la presente tesis se han estudiado diferentes características de las Flavobacterias, la clase más abundante entre los Bacteroidetes marinos, centrándonos sobre todo en dos rasgos principales: respuesta a la luz y la degradación de polímeros. Como representante modelo de las Flavobacterias escogimos *Polaribacter* sp. MED152. El género *Polaribacter* es uno de los más abundantes en el océano. En 2004, el 9% del total de secuencias del gen del 16S depositadas en GenBank de Bacteroidetes marinos pertenecía al género *Polaribacter* (Pommier *et al*., 2005), y de éstas, el 36% estaban estrechamente relacionadas con MED152. También se ha observado que MED152 es el décimo aislado que recupera más secuencias en estudios de metagenomas marinos (Yooseph *et al*., 2010). Además, usando FISH se ha encontrado que el género *Polaribacter* una muestra una abundancia media del 17% del total de células contadas en el océano Ártico (Nikrad *et al*., 2012). Por lo tanto, su elevada abundancia en el bacterioplancton marino junto con el hecho de poseer la proteorodopsina hacían de MED152 un genoma muy atractivo para ser estudiado.

Así, en el **Capítulo 1**, el estudio del genoma de MED152 abrió una ventana a la comprensión de la adaptación evolutiva de estos organismos para crecer en entornos oligotróficos iluminados. Con el fin de comprobar si se podrían encontrar patrones comunes en otras Flavobacterias y compararlos con otros grupos de bacterias, se analizaron genomas adicionales de Flavobacterias en los **Capítulos 2** y **3**, incluyendo algunos representantes de los grupos bacterianos más abundantes del océano como las Proteobacterias y las Cianobacterias. Se examinaron posibles patrones debido a la presencia del gen de la PR, en la degradación polímeros y en el contenido de islas genómicas. Se intentó comprobar experimentalmente algunas de las hipótesis derivadas de estos estudios *in silico* en los **Capítulos 4** y **5** por medio de cultivos puros y pruebas bioquímicas en las comunidades microbianas naturales. La Figura 11 muestra la relación entre los capítulos.

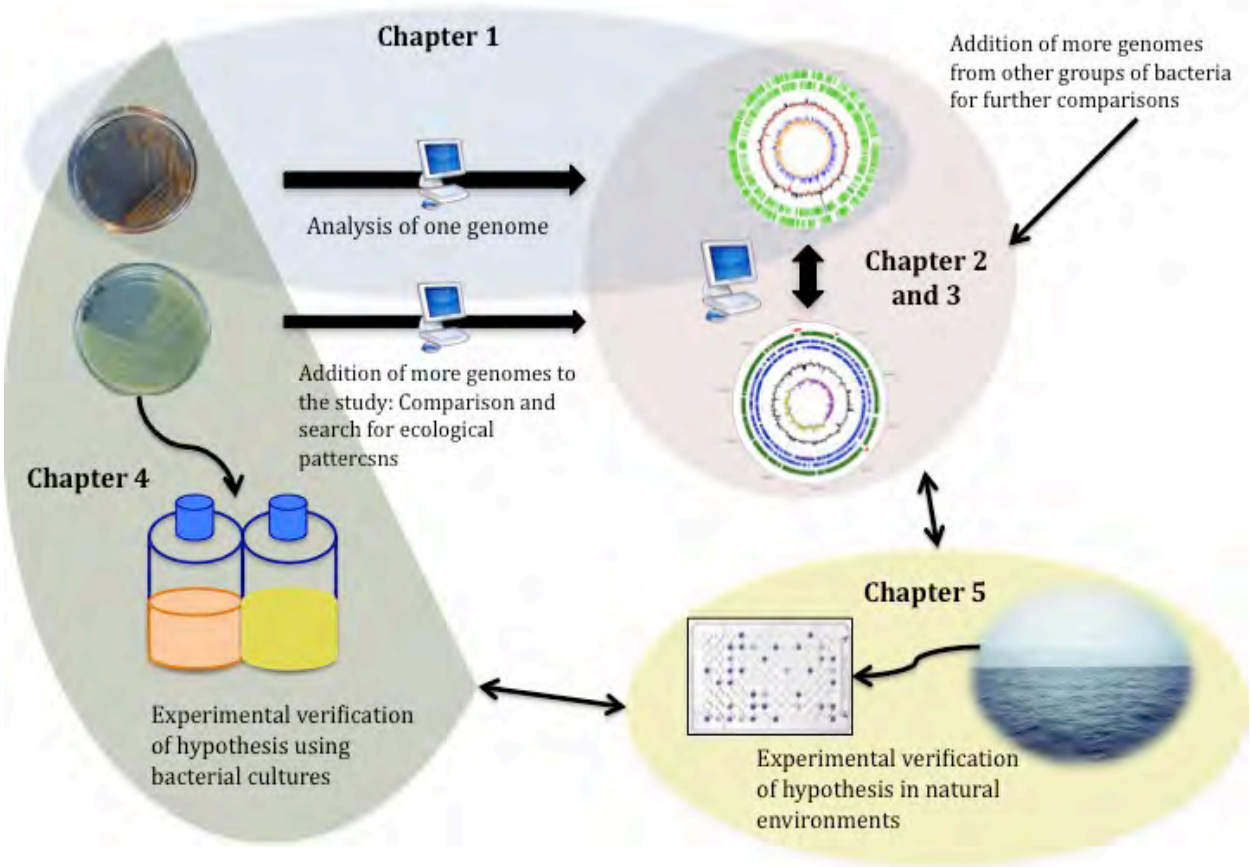


Figura 11. Relación entre los capítulos de esta tesis. El genoma *Polaribacter* sp. MED152 fue analizado en el **Capítulo 1**. En el **Capítulo 2** se compararon más Flavobacterias buscando patrones comunes que podrían estar presentes en sus genomas y en el **Capítulo 3** se añadieron todavía más genomas para el estudio de los patrones en las islas genómicas en el bacterioplancton marino. Como algunas de estas Flavobacterias habían sido aisladas, en el **Capítulo 4** las pudimos utilizar para llevar a cabo experimentos *in vivo* tratando de resolver algunas de las preguntas que aparecieron cuando analizamos sus genomas. Por último, en el **Capítulo 5** nos fuimos directamente al medio natural y estudiamos la comunidad microbiana *in situ*.

Alternancia de estilos de vida para sobrevivir en la escasez

Es arriesgado hablar de la ecología de un organismo simplemente mediante el análisis de su genoma, pero los genes que se encuentran en un genoma pueden proporcionar una gran cantidad de información sobre las posibles estrategias de estos organismos. En el **Capítulo 1**, el genoma de *Polaribacter* sp. MED152 sugirió una posible

alternancia de estilos de vida de acuerdo a los niveles de materia orgánica en el océano. Cuando las partículas de alto peso molecular (HMW-DOM) fueran abundantes, MED152 podría colonizarlas y moverse sobre su superficie por reptación. Pero una vez que todos los sustratos poliméricos hubieran sido consumidos, MED152 podría llevar una existencia de vida libre gracias a su capacidad para responder a la luz. Por otra parte, un aumento en la fijación de CO_2 en presencia de luz sugiere que su metabolismo central limitado podría complementarse con la fijación anaplerótica de carbono inorgánico.

a) La vida en abundancia: adhesión y degradación de los polímeros

En general, los Bacteroidetes son conocidos como consumidores eficientes de los polímeros tales como polisacáridos y proteínas (Cottrell y Kirchman 2000;. Woyke *et al.,* 2009). Siendo así, deberíamos encontrar en el genoma de *Polaribacter* evidencias que corroborasen este hecho. Y así fue, MED152 poseía un gran potencial glicolítico, debido a sus 29 glicosil hidrolasas y 81 peptidasas. También es posible que MED152 pudiera producir exopolisacáridos involucrados en la fijación no específica a partículas, ya que su genoma codifica 32 glicosil transferasas (Corpe *et al*., 1976; Sutherland *et al*., 1985).

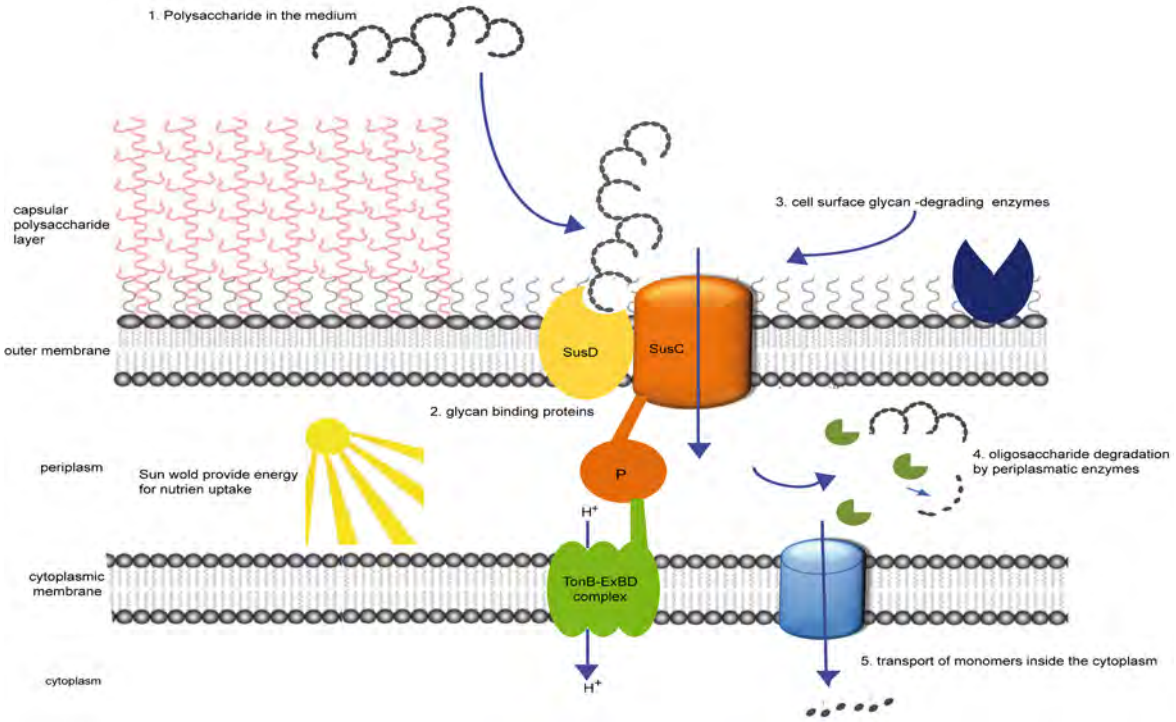


Figura 12. Modelo simplificado para el procesado los polisacáridos basado en el Sistema Sus en *Bacteroides thetaiotaomicron*. Adaptado de Martens EC *et al.,* J. Biol Chem (2009).

Se sabe también que la materia orgánica polimérica puede estar unida por complejos de membrana externa (Figura 12), formados por SusC, un canal de ligando, y SusD, una proteína de membrana externa (complejos SusCD), que actuan como entidades de unión a polisacáridos antes de iniciar su degradación (Anderson y Salyers 1989; Shipman *et al.,* 2000; Blanvillain *et al*., 2007). En estos complejos, las enzimas hidrolíticas se encuentran en la superficie de la bacteria, pero también hay proteínas de membrana externa que unen los polisacáridos a la superficie bacteriana (SusD) (Reeves *et al*., 1997). Se encontraron cinco complejos SusCD en MED152 (**Capítulo 1**, Fig. S7; **Capítulo 2**, Tabla SM9). Esta es una clara evidencia de que la adherencia y la degradación están estrechamente unidas a la utilización eficiente de polímeros.

Estudios recientes han encontrado que los transportadores dependientes de TonB eran muy abundantes en las comunidades microbianas marinas y, más aún, que la energía para el transporte de solutos podría estar proporcionada por la PR (Morris *et al*., 2010). De hecho, los transportadores dependientes de TonB son, junto con los transportadores ABC, el tipo de transportadores más abundante en MED152, representando 3,9% del total de transportadores.

Por otra parte, MED152 tiene un conjunto completo de genes (15) que participan en la motilidad por reptación, lo que sería beneficioso para la exploración de superficies sólidas. A pesar de que no se comprende totalmente, parece que la capacidad de deslizarse sobre la superficie podría proporcionar a las células una ventaja selectiva para la utilización de polímeros. Este es el mecanismo predicho para *C. hutchinsonii* (Xie *et al*., 2007) y *F. johnsoniae* (Braun *et al*., 2005) para la utilización de celulosa y quitina, respectivamente.

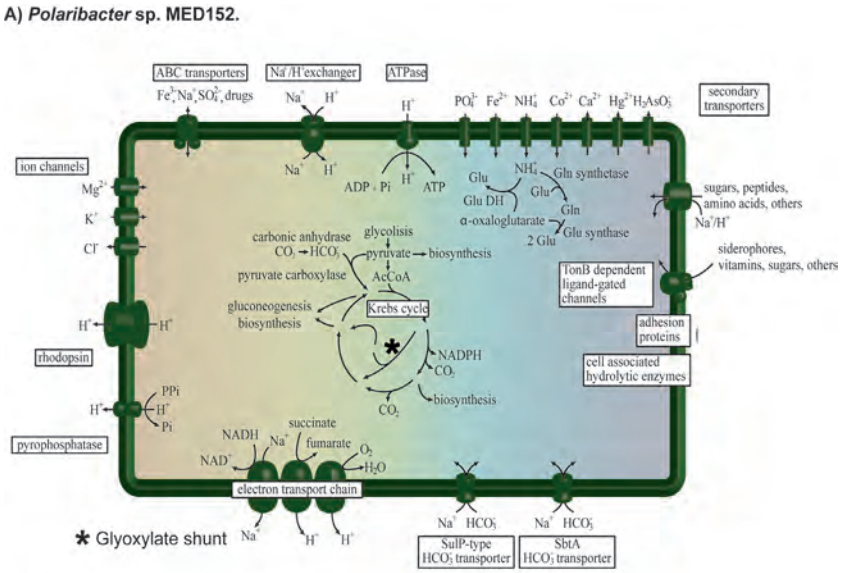
Además, hasta 26 genes contienen dominios involucrados en la adherencia a superficies (**Capítulo 1**, Tabla SM9), algunos de ellos en complejos hidrolíticos que incluyen SusC y SusD (**Capítulo 1**, Fig. SM7).

b) Cuando se agota el DOM: sistema de captación de la luz

La luz en un factor ambiental importante y la principal fuente de energía en la biosfera. Por tanto, los organimos han evolucionado para responder a ella para optimizar su crecimiento y metabolismo. Uno de los puntos clave del genoma de MED152 es la presencia del gen de la proteorodopsina (PR) y genes para la síntesis del retinal (cromóforo de la PR). Además contiene bastantes genes relacionados con respuestas de regulación a la luz (**Capítulo 1**, Tabla SM6). Fue interesante descubrir que MED152 posee un fitocromo, una proteína que responde a las regiones del rojo o rojo lejano del espectro de luz visible que también está presente en cianobacterias y en plantas, donde la activación del fitocromo regula aproximadamente el 10% de los genes presentes en plantas (Sharrok, 2008) . A pesar de que la vía completa de transducción de señales de la respuesta del fitocromo aún no ha sido descrita, existe evidencia de que en las plantas los fitocromos tienen actividad en la transducción de señales (Sharrok, 2008). Por lo tanto, su presencia en MED152 sugiere que esta proteína contribuye a la regulación de la síntesis de proteínas que responden a la luz.

La fototrofía mediada por la PR había sido sugerida para complementar las necesidades energéticas de las células en presencia de luz, disminuyendo la cantidad de carbono respirado. Sin embargo, esto representa un desafío para la célula, resultando en un desequilibrio del ciclo de Krebs, que MED152 podría resolver con la fijación anapleróticade CO_2, generando oxaloacetato a partir de bicarbonato a través de las enzimas PEP y piruvato carboxilasas (**Capítulo 1**, Fig. SM4). Cuando la luz es la única fuente de ATP, el ciclo de Krebs es supuestamente utilizado para fines biosintéticos y no para la generación de energía (como en los organismos autótrofos). El ciclo del glioxilato (que también está presente en MED152, **Capítulo 1**, Fig. SM3) permite el crecimiento con algunos compuestos que entran en la parte derecha del ciclo de Krebs, como lípidos o acetato. Pero en MED152, que puede obtener energía de la luz, esta vía metabólica podría jugar un papel importante en la regeneración de oxalacetato. La vía alternativa del glioxilato no está presente en las bacterias autotróficas, pero sí en plantas y muchas otras bacterias.

La Figura 13 compara *Prochlorococcus marinus* y *Polaribacter* sp. MED152. Se observa el tipo de ciclo de Krebs en herradura en *Prochlorococcus*. Esto es típico de autótrofos estrictos, ya que no necesitan completar el ciclo por no ser oxidativo. Por lo tanto carecen de una de las enzimas: la alfa-cetoglutarato deshidrogenasa.



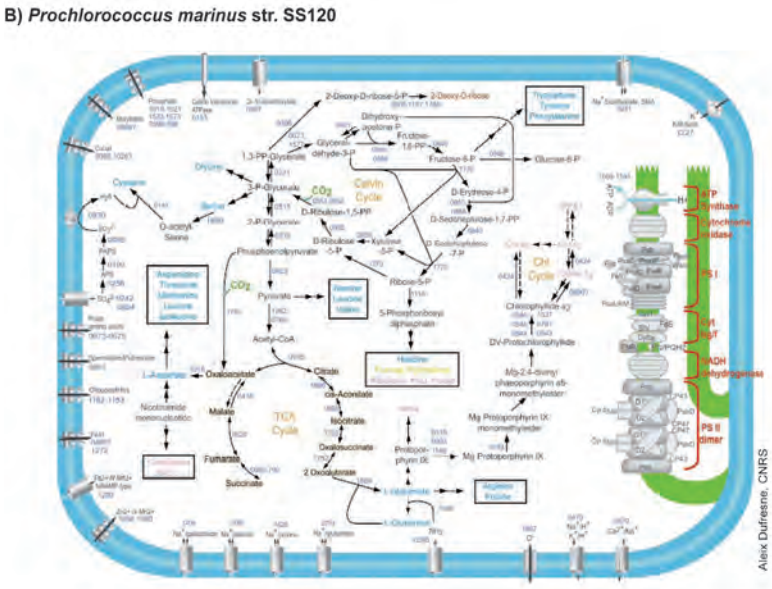


Figura 13. Comparación entre (A) *Polaribacter* sp. MED152 y (B) *Prochlorococcus marinus* str. SS120. Nótese que los transportadores son similares porque incorporan los mismos nutrientes y también se basan en el gradiente de Na^+. Algunas diferencias importantes son el ciclo de Calvin y la parte derecha de la célula donde aparecen los fotosistemas y la cadena de transporte de electrones. MED152 no tiene el ciclo de Calvin ni cualquier otro ciclo para fijar CO_2 autotróficamente.

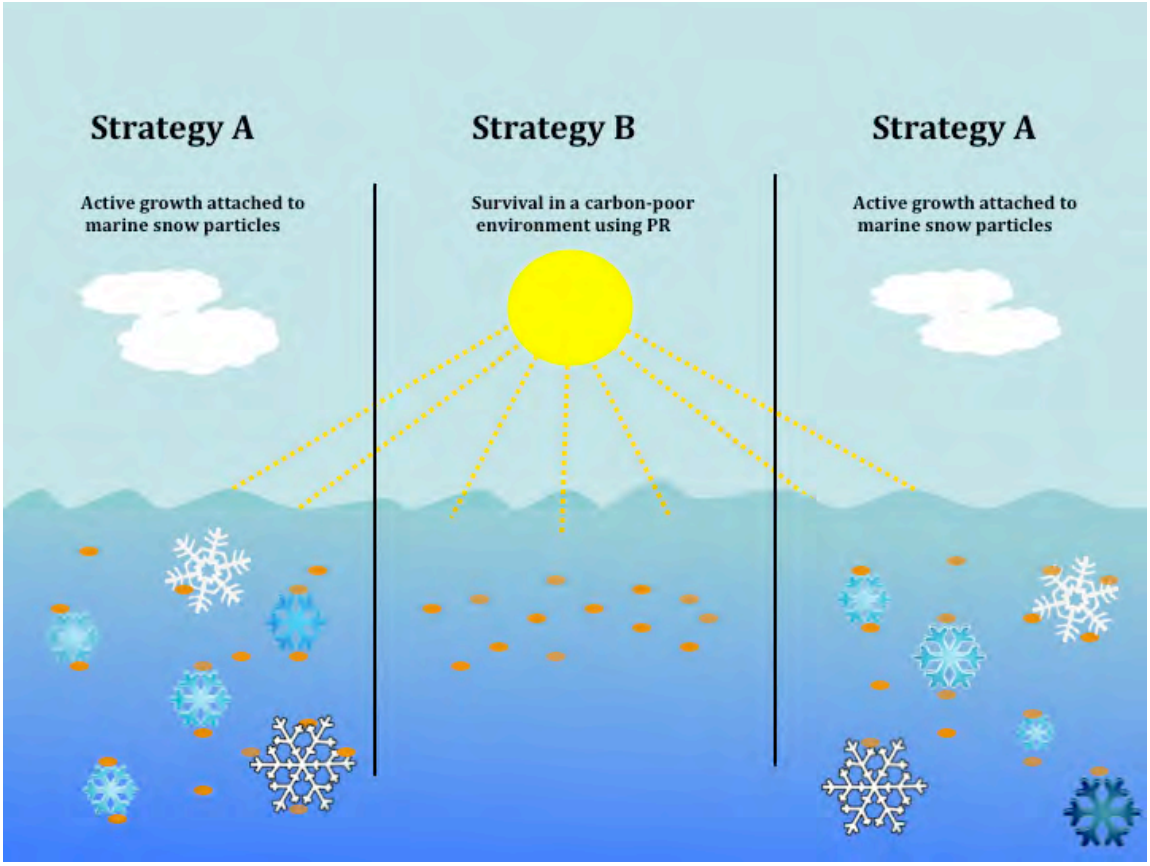


Figura 14. Modelo sobre las estratégias de vida de *Polaribacter* sp. MED152 en el océano.

Encontramos que MED152 fija mayor cantidad de bicarbonato en la luz que en la oscuridad cuando crece en un medio relativamente rico (Marine Broth® 8 veces diluido) (**Capítulo 1**, Fig. 3 y **Capítulo 4**, Fig. 7). Esto sugiere que la fijación anaplerótica de carbono inorgánico estimulada por la luz podría permitir a las Flavobacterias que contienen proteorhodopsina utilizar eficientemente la materia orgánica para biosíntesis.

Por lo tanto, teniendo la capacidad de unirse a partículas, degradarlas e incorporar las moléculas pequeñas resultantes al interior de la célula, MED152 sería capaz de sobrevivir mientras haya materia orgánica presente en el océano. Una vez que estas partículas se han agotado, la PR podría proporcionar la energía necesaria para sobrevivir sin necesidad de respirar compuestos orgánicos. La Figura 14 muestra un modelo hipotético de las estrategias de MED152 en el océano.

Patrones entre Bacteroidetes marinos

El primer genoma íntegro de Bacteroidetes marino que se analizó fue el de *Gramella forsetii* KT0803 (sin PR) (Bauer *et al*., 2006), indicando una clara adaptación a la degradación de compuestos de alto peso molecular. *Polaribacter* sp. MED152 fue el segundo genoma de esta clase taxonómica que se analizó y resultó tener las mismas adaptaciones para la degradación de polímeros (**Capítulo 1**). Dado que esta habilidad de degradación parece ser un patrón recurrente e independiente al hecho de poseer la PR, la comparación con más genomas de Bacteroidetes marinos, con (PR +) y sin (PR-) proteorodopsina, nos ayudaría a identificar los patrones generales entre representantes de esta clase taxonómica.

a) Correlación entre la posesión del gen PR y el tamaño del genoma.

Es curioso que todos los genomas de Bacteroidetes PR+ son pequeños (**Capítulo 2**, Fig. 1). Como se vio en el **Capítulo 1**, el genoma de MED152 es muy pequeño (2,9 Mb). MED134 también posee el gen de la PR y los dos tienen genomas más pequeños que los otros dos genomas PR-, *Gramella forsetii* y MED217 (**Capítulo 2**, Tabla 1). Es tentador pensar que de alguna manera la PR permite a las bacterias reducir sus genomas, de modo que no sería necesario poseer los genes de muchas vías metabólicas diferentes para la utilización de carbono porque la energía de la luz haría que fuesen más independientes de compuestos orgánicos como fuente de energía.

b) ¿Están las bacterias PR+ mejor adaptadas a vivir en condiciones oligotróficas?

En el **Capítulo 1** propusimos que la PR, junto con otros genes de detección a la luz y fijación anaplerótica de carbono, representaban adaptaciones para sobrevivir en la superficie del océano cuando las concentraciones de materia orgánica eran bajas. El siguiente paso fue probar si la capacidad de responder a la luz y la existencia de la fijación de CO_2 en otras Flavobacterias eran características compatibles con la presencia o ausencia de la PR. En cuanto a los dominios de respuesta a la luz, no se observaron diferencias significativas en número al comparar bacterias PR + y PR- (**Capítulo 2**, Fig. 5). Sin embargo, con la incorporación de más grupos bacterianos para el análisis,

como cianobacterias o proteobacterias, aparecieron diferentes patrones de dominios de respuesta a la luz (**Capítulo 2**, Tabla 3), lo que sugiere una ligera diferencia en la estrategia de su uso. Respecto a las enzimas que intervienen en la fijación de CO_2 (comentadas en el **Capítulo 1**: piruvato y la PEP carboxilasa, isocitrato liasa, malato sintasa), hubo una clara correlación entre la posesión de PR y el número de genes antes mencionados (**Capítulo 2**, Fig. 6B).

c) ¿Siguen las bacterias PR + la misma estrategia que MED152 cuando están limitados de materia orgánica?

Las adaptaciones a la adhesión y degradación de partículas de MED152 fueron mencionadas en el **Capítulo 1**. Se observó que las cuatro Flavobacterias tenían un gran número de proteínas con dominios de adhesión a superficies (**Capítulo 2**, Tabla SM1). Esto también es cierto para otras Flavobacterias, sin embargo, este rasgo no aparecía en otros grupos de bacterias marinas (**Capítulo 2**, Fig. 8A y 8B). Esto parece una característica en común entre flavobacterias independientemente de la capacidad de responder a la luz, ya que no se observaron diferencias entre especies PR + y PR-.

El potencial de degradación observado en el **Capítulo 1** también fue corroborado con Flavobacterias adicionales añadidas al estudio (**Capítulo 2**, Fig. 8C y 8D) que, en general (con algunas excepciones) presentaban la mayor cantidad de glicosil hidrolasas y peptidasas por megabase en comparación con otros grupos. Además, los Bacteroidetes PR+ mostraban significativamente más peptidasas por megabase que los Bacteroidetes PR-, sugiriendo una especialización en la degradación de proteínas (**Capítulo 2**, Fig. 10) y todos ellos mostraron diferentes conjuntos de estas enzimas, lo que indica una especialización para determinados tipos de proteínas (**Capítulo 2**, Fig. 11). Así, cada Flavobacteria estaría adaptada a nichos ligeramente diferentes lo que haría dismunir la competencia entre ellos.

La Figura 15 muestra la cantidad de glicosil hidrolasas, peptidasas, glicosil transferasas y proteínas de adhesión en grupos como Bacteroidetes, Planctomycetes y Proteobacterias. Nuestras cuatro Flavobacterias tienen la mayor cantidad de glicosil hidrolasas por megabase (Mb), después de los dos *Bacteroides* spp. que son bien conocidos

como especialistas en la degradación de polisacáridos, y que albergan el mayor número de peptidasas por Mb después de la gammaproteobacteria *Idiomarina loihiensis* cuyo suministro de carbono y energía se basa principalmente en aminoácidos (Hou *et al*., 2004).

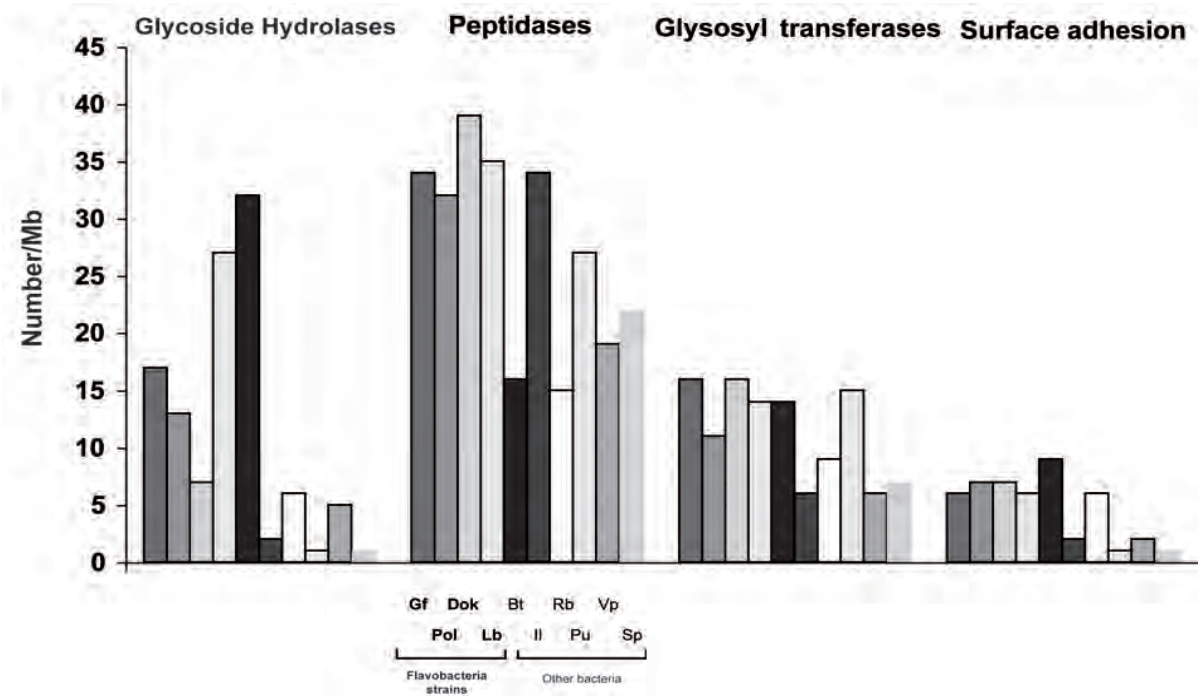


Figura 15. Comparación de las capacidades hidrolíticas y del potencial de adhesión de algunos miembros del bacterioplancton heterotrófico marino y de Bacteroidetes simbiontes del tracto digestivo de los humanos. Gf: *Gramella forsetii,* **Dok:** *Dokdonia* sp. MED134*,* **Pol:** *Polaribacter* sp. MED152, **Bt:** *Bacteroides thetaiotaomicron,* **Lb:** *Leeuwenhoekiella blandensis* MED217*,* **Rb:** *Rhodopirellula baltica,* **Il:** *Idiomarina lohiensis,* **Vp:** *Vibrio parahaemolyticus,* **Pu:** *Pelagibacter ubique,* **Sp***: Silicibacter pomeroyi.*

Cuantificación de islas genómicas (IG) en bacterias marinas

Las bacterias planctónicas viven y compiten en la zona fótica de los océanos bajo la influencia penetrante de la luz (Frigaard *et al*., 2006). La única manera de evitar la competencia es evolucionando y dando lugar a nuevas estrategias de vida. La transferencia horizontal de genes es uno de los mecanismos evolutivos que amplia el *pool* flexible del genoma en poblaciones bacterinas, lo que facilita la adaptación a nuevos nichos ecológicos (Doolittle, 1999;. Boucher *et al*., 2003). Los nuevos genes transferidos se agrupan en lo que se denominan islas genómicas (IGs). Recientemente, se ha observado que *Dokdonia*

sp. MED134 tiene los genes directamente involucrados en la fototrofía mediada por la PR (*prd*: PR y *blh*: ß-caroteno) y un sensor a la luz azul cerca de un transposón conjugativo (González *et al*., 2011). Estos genes podrían ser transferidos a otros genomas y, si albergan los genes para la síntesis de carotenoides, se podrían convertir fácilmente en fototrofas. Y al igual que la PR, muchos otros genes transferidos podrían proporcionar un beneficio potencial para aumentar la aptitud para competir ("*fitness*") de las bacterias. Y ese fue el siguiente paso en el estudio del potencial genético de los Bacteroidetes marinos. Para ello, en el **Capítulo 3** se construyó una base de datos que incluye 8152 genes presentes en las IG en 66 de los 70 genomas bacterianos seleccionados (**Capítulo 3**, Tabla SM1). En el **Capítulo 2**, vimos que las IGs eran una de las características que distinguen unas Flavobacterias de otras mediante el análisis de una IG de *L. blandensis* MED217, pero esta vez estudiamos los genomas de más flavobacterias y añadimos genomas de otros grupos de bacterias con el fin de compararlos. En conjunto, se detectaron 438 IGs, abarcando un total de 8,87 Mb (**Capítulo 3**, Tabla SM3). Desafortunadamente era imposible detectar las islas genómicas manualmente como se había hecho con algunas cepas de *Prochlorococcus* (Coleman *et al*., 2006) y *Synechococcus* (Drufresne *et al*., 2008) debido al gran número de genomas y porque algunos no estaban cerrados todavía. Por esta razón tuvimos que recurrir a herramientas de búsqueda automática y comprobamos que, aunque perdíamos alguna IG, podíamos estar seguros de que lo que detectábamos era realmente una IG. Las islas genómicas representaban el 0-12% de los genomas y en la mayoría de genomas mostraban un porcentaje que oscilaba entre el 2 y el 5% (**Capítulo 3**, Fig. 3).

Arquitectura y anotación funcional de las islas genómicas en bacterias marinas

Si la transferencia horizontal de genes (incluye transformación, conjugación y transducción) es el mecanismo descrito por el cual las bacterias adquieren nuevos genes, cabría esperar encontrar un gran número de elementos genéticos móbiles como transposasas, transposones conjugativos, integrasas o genes relacionados con integrasas de fagos (Dobrindt *et al*., 2004; Juhas *et al*., 2008). Y así fue: encontramos que el 70% de las islas descritas contenían estos elementos móbiles.

Tal y como esperábamos, comprobamos que la mayoría de las islas habían sido originadas a través de la transferencia horizontal de genes (**Capítulo 3**, Fig. 4), por

el hecho de tener un gran porcentage de proteínas hipotéticas y diferentes genes en diversos organismos. Sin embargo, para nuestra sorpresa, encontramos otra estructura de IG, transferidas a través de recombinación homóloga (**Capítulo 3,** Fig. 5). Este tipo de islas está caracterizada por tener genes del metabolismo central ("*housekeeping*"), un número bajo de proteínas hipotéticas y la misma estructura en diferentes organismos. Las Flavobacterias fueron uno de los grupos con mayor número de este tipo de islas.

Para poder demostrar el papel crucial de las IGs en la evolución de las bacterias es necesario saber la función de los genes contenidos en las islas. Alrededor de un 55-60% de los genes no pudieron ser asignados a ninguna de las categorías funcionales, por lo tanto, estas proteínas hipotéticas podrían considerarse como nuevas. El resto de genes fueron anotados con el fin de asignarlos a una categoría funcional usando dos aproximaciones (**Capítulo 3**, Fig. 8 y Fig. 9) y revelaron una amplia gama de genes biológicamente relevantes. Además, se sabe que los diferentes grupos de bacterias marinas podrían tener diferentes estrategias para vivir en el océano. El hecho de encontrar diferencias en las categorías funcionales representadas en nuestros cuatro grupos de estudio revelaría el porqué de sus adaptaciones al medio marino. Así, la comparación de genes con categorías funcionales asignadas (**Capítulo 3**, Tabla 2 y Tabla SM4) mostró diferencias significativas en todos los casos, sugiriendo que cada grupo filogenético podría tener una estrategia diferente en el papel desempeñado por las IGs en su ecología. Es probable que muchos genes dentro de las IGs esten seleccionados positivamente debido a los potenciales beneficios de estos genes en el estilo de vida de las bacterias.

Más allá de los genomas: analizando la función fisiológica de PR

Desde que el gen de la PR fue descubierto (Béjà *et al*., 2000), se ha comprobado su presencia en varios grupos de bacterias marinas, a pesar de que todavía no se ha descrito una función fisiológica consistente para este gen. Tres capítulos de esta tesis (**Capítulos 1, 2** y **4**) se han centrado en las implicaciones/beneficios de poseer la PR. Los **Capítulos 1** y **2** trataron de hipotetizar su papel en el medio natural a partir de lo que "está escrito" en los genomas bacterianos. En el **Capítulo 4** se trató de descifrar la función fisiológica de esta proteína por medio de experimentos de crecimiento con dos aislados que poseen la PR.

Como se ejemplifica en la figura 16, los experimentos de cultivo de agua de mar con organismos que han sido previamente aislados son necesarios para comprender la fisiología y la importancia potencial en el entorno de la PR. Experimentos con agua de mar artificial mostraron diferentes respuestas del crecimiento en varias condiciones.

En el **Capítulo 1** hipotetizamos la posible función de la PR en MED152. Basándonos en algunos de los resultados obtenidos en el **Capítulo 2**, ya se mencionó que había evidencias en el contenido de los diferentes genomas que sugería posibles formas de utilización de la PR. Los resultados obtenidos en el **Capítulo 4** también podrían indicar un uso diferenciado de la PR en MED152 y MED134, probablemente con el fin de adaptarse a nichos ligeramente diferentes. Mientras MED134 mostró mayores diferencias de rendimiento entre las incubaciones a la luz y a la oscuridad (Gómez-Consarnau *et al.*, 2007) que MED152 (**Capítulo 4**, Fig. 6.); MED152 mostró mayores tasas de fijación de carbono (**Capítulo 4**, Fig. 7A). Fue interesante observar que las diferencias en la respuesta a la luz, además de producirse a bajas concentraciones de materia orgánica,

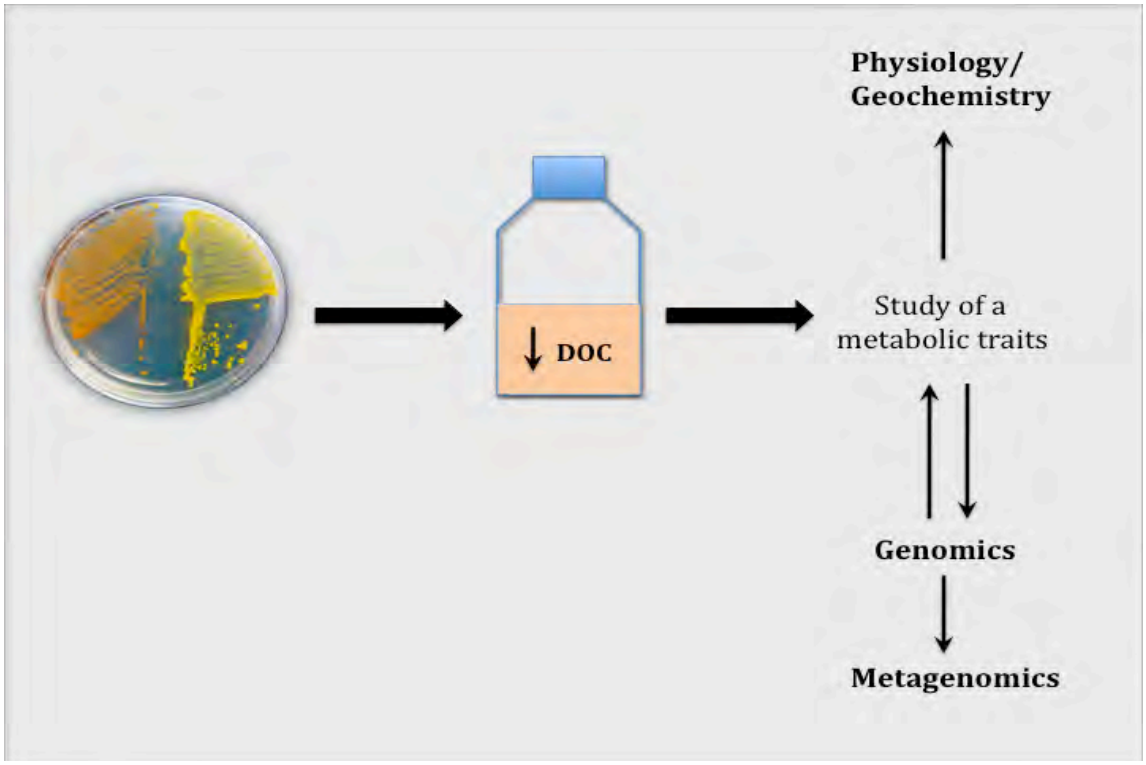


Figura 16. Representacion esquemática de la importancia de los cultivos puros para la deterterminación de funciones bacterianas.

también dependendían de la composición de esta materia orgánica (**Capítulo 4**, Fig. 4-6). Desafortunadamente, nuestros resultados no aclaran esta observación y, por tanto, son necesarios más experimentos con diferentes medios y condiciones de crecimiento a fin de demostrar los efectos reales de la PR .

Capacidades metabóbicas de la comunidad microbiana en el Ártico

Volviendo al potencial hidrolítico de los Bacteroidetes marinos discutido ya en los **Capítulos 1** y **2**, en el **Capítulo 5**, la capacidad de degradación de polímeros (y otros substratos) fue analizada en el medio natural (aprovechando que surgió la oportunidad de una campaña oceanográfica durante el Año Polar Internacional), usando placas Biolog como técnica de identificación metabólica. Ya se había visto en estudios anteriores en el Ártico (2003-2004) que las placas de Biolog Ecoplate® habían sido útiles para determinar las capacidades metabólicas de las comunidades microbianas. Para el **Capítulo 5** desarrollamos unas placas de Biolog personalizadas, las llamadas MT2®, con la intención de analizar la degradación de distintos compuestos por la comunidad natural (**Capítulo 5**, Tabla 3) a diferentes profundidades y a lo largo del año (de febrero a julio).

A diferencia de las Ecoplate®, las MT2® son placas cuyos pozuelos carecen de fuentes materia orgánica y tienen que ser suplementadas con los diferentes compuestos a estudiar, pero el primer problema que se plantea es la determinación de la mejor concentración a la cual tienen que estar los compuestos con el fin de detectar su oxidación. Fue difícil de decidir, ya que la concentración de los compuestos en las placas Eco (algunos de los cuales añadimos en las placas MT2), nuestro referente, no ha sido todavía facilitada por el fabricante. Además, no todos los compuestos reaccionan de la misma manera a una cierta concentración, mientras unos pueden ser solubles en agua, otros (especialmente polímeros) pueden precipitar. Además, el estudio de la diversidad funcional medida a partir de la utilización de compuestos de carbono usando las placas MT2 había sido llevada a cabo mayoritarimente en comunidades microbianas de suelos, en algunos casos de agua dulce y solo en una ocasión fue utilizada para determinar el potencial de utilización de compuestos por bacterias de ambientes marinos. Así, de los 23 sustratos probados, sólo 17 fueron utilizados (**Capítulo 5**, Fig. 2, Tabla 2), pero no se descarta que el resto de los compuestos no fueron utilizados, simplemente no se

observó su oxidación, quizás porque la actividad bacteriana no fue suficiente o porque la concentración del compuesto era demasiado baja.

Observamos que en superficie las bacterias utilizaron más sustratos que en profundidad (**Capítulo 5**, Fig. 2 y Fig. 3) y más frecuentemente, aunque no significativamente para ningún compuesto, excepto para el conjunto de aminoácidos. Con respecto a la época del año, no hubo actividad detectable durante el invierno, pero el número de casos positivos comenzó a aumentar con el aumento de clorofila *a* en el agua, desde mediados de marzo en adelante, a excepción de un período de 4 semanas cuando la señal desapareció por completo (**Capítulo 5,** Fig. 3). Este "agujero" que se encuentra en medio de un período de alta actividad podría deberse a un cambio en la comunidad de bacterioplancton (observado en una DGGE) debido a la transición de las condiciones de invierno a las de primavera, cuando la comunidad adaptada al invierno desaparecía y la adaptada a las condiciones de primavera/verano aparecía (**Capítulo 5**, Fig. 1SM).

Los resultados se compararon con los de un estudio previo (campaña CASES, 2003-2004) llevado a cabo en la misma región, pero en otra área hidrográfica, la Bahía de Franklin. En este último estudio, los carbohidratos poliméricos produjeron el mayor número de resultados positivos, al igual que en CFL, pero los aminoácidos mostraron una proporción mucho menor de resultados positivos (**Capítulo 5**, Fig. 5). La Figura 17 muestra la diferencia en el compuesto más utilizado durante el CFL (A) y en CASES (B). La proporción de hidratos de carbono y polímeros utilizados fueron mayores en CASES y la proporción de aminoácidos fue mayor en CFL. En ambos estudios la actividad en la superficie fue mayor que en profundidad.

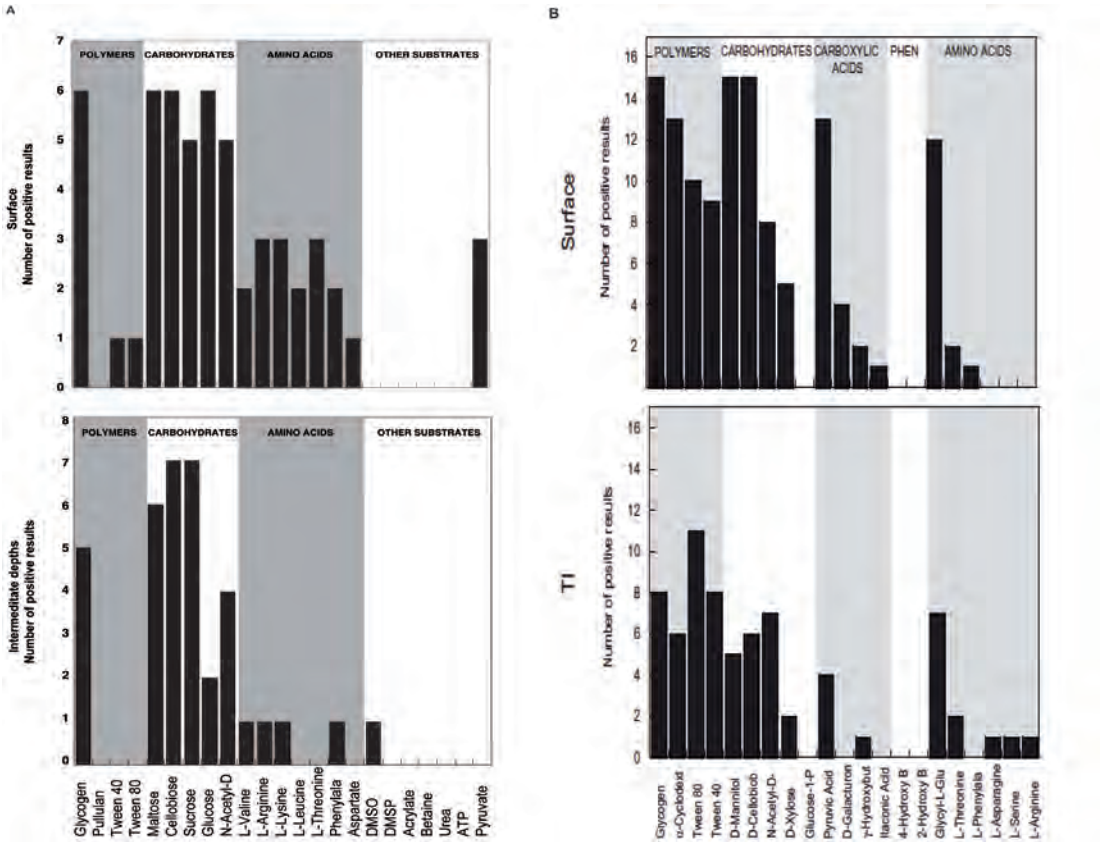


Figura 17. Número de resultados positivos para los sustratos en microplacas Biolog MT2 durante CFL (A) y en microplacas Eco durante CASES (B).

Otra diferencia con CASES fue la época del año con más actividad. Mientras que la parte más activa del año en CFL fue a partir de la primavera, siguiendo el mismo patrón que la clorofila *a*, en CASES el invierno fue el período del año más activo y la disminución de la actividad comenzó la primavera (Fig. 18). Es difícil dar una respuesta a la posible causa de estas diferencias observadas entre los dos estudios, pero una posible razón se deba probablemente a las diferentes condiciones hidrográficas que presentan la Bahía de Franklin y el Golfo de Amundsen.

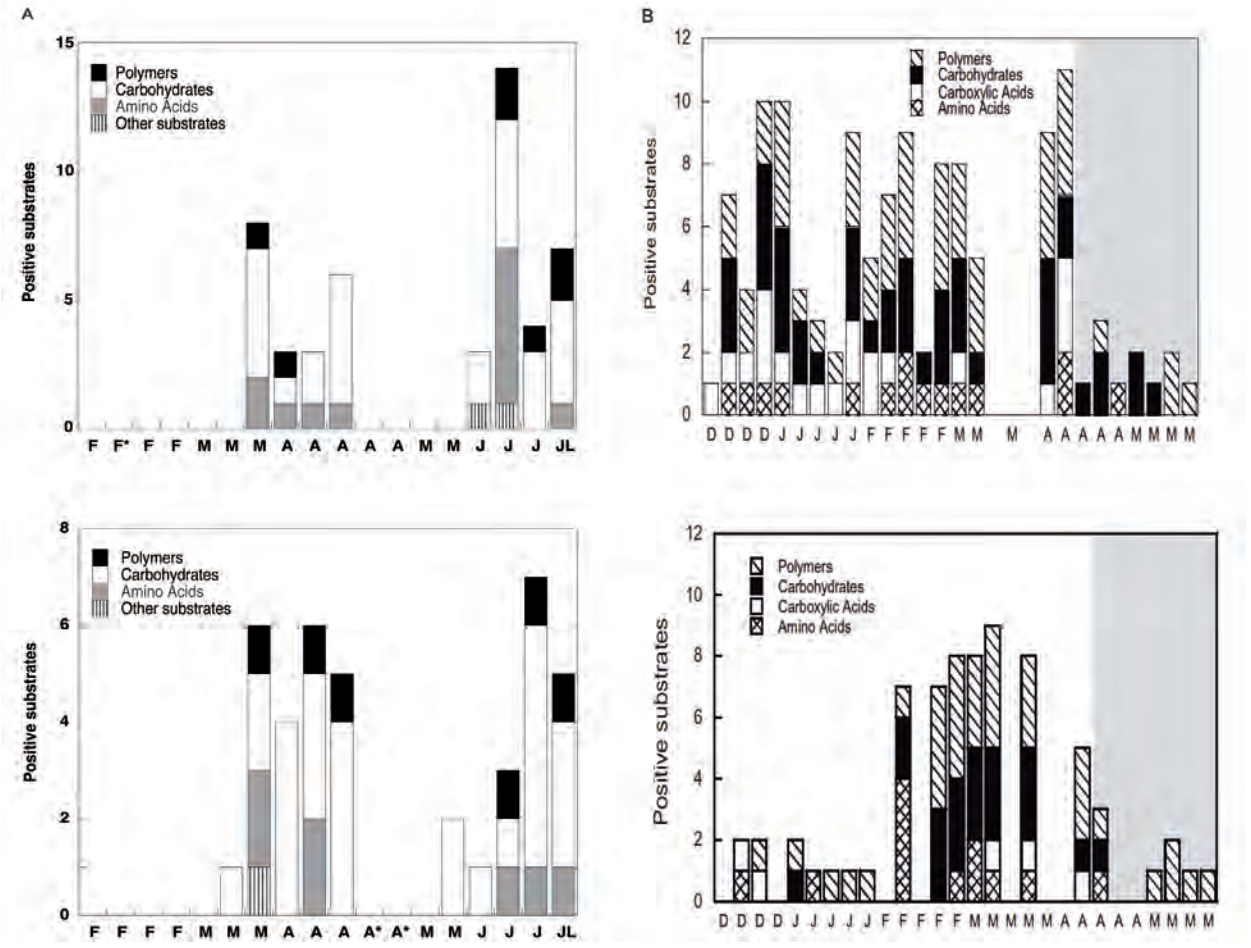


Figura 18. Número de sustratos usados a lo largo del año, separados en categorias para CFL (A) y CASES (B).

CONCLUSIONES

1. Las características generales del genoma de *Polaribacter* sp. MED152 son consistentes con una vida en la superficie del océano, incluida la utilización de la luz y un número significativo de proteínas dependientes de Na^+.

2. Características más específicas de *Polaribacter* sp. MED152 revelan adaptaciones para alternar entre la adhesión a partículas y la vida libre.

3. Utilizando la proteorodopsina, *Polaribacter* sp.MED152 es capaz de capturar la energía que proviene de la luz, aumentar la fijación anaplerótica de bicarbonato y probablemente utilizará los pocos compuestos de carbono que puede conseguir exclusivamente para la biosíntesis.

4. Los Bacteroidetes marinos comparten la capacidad de adhesión, motilidad por repatación, y un gran número de enzimas degradadoras de polímeros.

5. Los Bacteroidetes marinos que contienen PR tienen genomas especialmente pequeños, con más genes implicados en la fijación de CO_2 por Mb que los Bacteroidetes sin PR.

6. Las islas genómicas son una característica común en la mayoría de los genomas bacterianos marinos examinados. Cada grupo de bacterias marinas posee un conjunto diferente de genes, indicando posibles diferentes estrategias en el papel que juegan las islas genómicas en la ecología de cada grupo.

7. La transferencia horizontal de genes y la recombinación homóloga tienen un papel importante en la movilidad de fragmentos de DNA entre diferentes taxones y en genomas estrechamente relacionados, respectivamente, modulando el “pool” flexible del genoma.

8. La anotación funcional de los genes presentes en las islas genómicas reveló una ámplia gama de genes biológicamente relevantes .

9. La luz estimula el crecimiento *Polaribacter* sp. MED152, Flavobacteria que contiene la proteorodopsina, en ambientes limitados por la materia orgánica.

10. La composición del medio de cultivo influye en la probabilidad de encontrar diferencias de crecimiento entre condiciones de luz y oscuridad.

11. La proteorodopsina incrementa la incorporación de bicarbonato en *Polaribacter* sp. MED152 cuando crece en diferentes medios pobres.

12. Los hidratos de carbono, particularmente moléculas poliméricas, son una fuente importante de carbono para el bacterioplancton ártico. Los aminoácidos también son consumidos, pero su incorporación es bastante variable dependiendo de las áreas y la época del año.

PERSPECTIVAS DE FUTURO

El interés por las Flavobacterias está aumentando cada día, junto con el número de microoorganismos de este grupo que estan siendo aislados en cultivo puro y los genomas que estan siendo secuenciados. Basándonos en los resultados obtenidos en esta tesis, el examen de algunos de los siguientes aspectos aumentaría la comprensión del papel de los microorganismos (en nuestro caso los Bacteroidetes marinos) en el océano:

1. Secuenciación completa y anotación de más miembros de los Bacteroidetes marinos y no marinos. Aunque el número de aislados es alta, para su estudio sería necesario tener sus genomas completos y todos los genes anotados con el fin de entender y compararlos con otros representantes del mismo o diferentes grupos.

2. Se debe hacer un gran esfuerzo para tratar de descubrir las funciones de las proteínas hipotéticas, que siempre representan un alto porcentaje del genoma y, probablemente, estas proteínas contienen la esencia que hace a cada organismo único.

3. Mejora y desarrollo de técnicas bioinformáticas para el análisis y comparación de varios genomas a la vez. Cada día, hay más y más información para procesar y cada vez más se necesitan máquinas más sofisticadas y poderosas para hacer frente a miles de gigabytes de datos. Además, se requerirían nuevos porgramas de ordenador que permitiesen análisis eficientes en un tiempo relativamente corto.

4. Se necesitan más esfuerzos para desarrollar un medio de cultivo completamente definido para *Polaribacter* y *Dokdonia*. Asimismo, serían necesarios más estudios para determinar por qué la luz produce diferencias en el crecimiento de estas bacterias con ciertas fuentes de carbono y no con otras.

5. Por último, se deben diseñar placas BIOLOG nuevas, más específicas y probando con nuevos sustratos no utilizados hasta ahora y que pudieran aumentar nuestra comprensión de la diversidad metabólica del bacterioplancton heterotrófico.

References

(Includes Introduction, Discussion and Spanish Summary)

& Annexes

REFERENCES (includes Introduction, Discussion, and Spanish Summary)

Alldredge, A.L., Cole, J.J. and Caron, D.A., 1986. Production of feterotrophic Bacteria inhabiting macroscopic organic aggregates (marine snow) from surface waters. Limnol Oceanogr, 31(1): 68-78.

Alonso, C., Warnecke, F., Amann, R. and Pernthaler, J., 2007. High local and global diversity of Flavobacteria in marine plankton. Environ Microbiol, 9(5): 1253-1266.

Alonso-Sáez, L. and Gasol, J., 2007. Seasonal variations in the contributions of different bacterial groups to the uptake of low-molecular-weight compounds in Northwestern mediterranean coastal waters. Appl Environ Microbiol, 73(11): 3528-3535.

Anderson, K.L. and Salyers, A.A., 1989. Genetic evidence that outer membrane binding of starch is required for starch utilization by *Bacteroides thetaiotaomicron*. J Bacteriol, 171(6): 3199-3204.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990. Basic local alignment search tool. J Mol Biol, 215(3): 403-410.

Amaral-Zettler, L., Artigas, L.F., Baross, J., Loka Bharathi, P.A., Boetius, A., Chandramohan, D., *et al*., 2010. A Global Census of Marine Life, in: Alasdair D. McIntyre, Ed., Life in the World's Oceans. Blackwell Publishing Ltd., 223-245 pp.

Amon, R.M.W. and Benner, R., 1994. Rapid cycling of high-molecular-weight dissolved organic matter in the ocean. Nature, 369(6481): 549-552.

Amon, R.M.W. and Benner, R., 1996. Bacterial utilization of different size classes of dissolved organic matter. Limnol Oceanogr, 41(1): 41-51.

Arnosti, C., Durkin, S. and Jeffrey, W.H., 2005. Patterns of extracellular enzyme activities among pelagic marine microbial communities: implications for cycling of dissolved organic carbon. Aquat Microb Ecol, 38(2): 135-145.

Azam, F., 1998. Microbial control of oceanic carbon flux: The plot thickens. Science, 280(5364):694-696

Azam, F., Fenchel, T., Field, J. G., Gray, J. S. and Meyer-Reil, L. A., 1983. The ecological role of water-column microbes in the sea. Mar Ecol Prog Ser, 10: 257-263.

Bauer, J.E., Williams, P. M. and Druffel, E. R. M., 1992. 14C activity of dissolved organic carbon fractions in the north-central Pacific and Sargasso Sea. Nature, 357(6380): 667-670.

Bauer, J.E., 2002. Carbon isotopic composition of DOM, in: Dennis A. Hansell and Craig A. Carlson Eds., Biogeochemistry of Marine Dissolved Organic Matter. Academic Press, San Diego, 405-453 pp.

Bauer, M., Kube, M., Teeling, H., Richter, M., Lombardot, T., Allers, E., *et al*., 2006. Whole genome analysis of the marine BacteroidetesGramella forsetii reveals adaptations to

degradation of polymeric organic matter. Environmental Microbiology, 8(12): 2201-2213.

Béjà, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L. *et al*., 2000. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science, 289(5486): 1902-1906.

Benner, R., 2002. Chemical composition and reactivity, in: Dennis A. Hansell and Craig A. Carlson Eds., Biochemistry of Marine Dissolved Organic Matter. Academic Press, San Diego, 59-90 pp.

Blanvillain, S., Meyer, D., Boulanger, A., Lautier, M., Guynet, C., Denancé, N. *et al* ., 2007. Plant carbohydrate scavenging through TonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. PLoS ONE, 2.

Boucher, Y., Douady, C.J., Papke, R.T., Walsh, D. A., Boudreau, M.E.R., Nesbø, C. L. *et al*., 2003. Lateral gene transfer and the origins of prokaryotic groups. Ann Rev Genet, 37(1): 283-328.

Braun, T.F., Khubbar, M.K., Saffarini, D.A. and McBride, M.J., 2005. *Flavobacterium johnsoniae* gliding motility genes identified by *mariner* Mutagenesis. J Bacteriol, 187(20): 6943-6952.

Bryden, H.L., Longworth, H. R. and Cunningham, S. A., 2005. Slowing of the Atlantic meridional overturning circulation at 25° N. Nature, 438(7068): 655-657.

Campbell, B.J., Waidner, L.A., Cottrell, M.T. and Kirchman, D.L., 2008. Abundant proteorhodopsin genes in the North Atlantic Ocean. Environ Microbiol, 10(1): 99-109.

Chen, K. and Pachter, L., 2005. Bioinformatics for whole-genome shotgun sequencing of microbial communities. PLoS Comput Biol, 1(2): e24.

Chin, W.-C., Orellana, M.V. and Verdugo, P., 1998. Spontaneous assembly of marine dissolved organic matter into polymer gels. Nature, 391(6667): 568-572.

Chisholm, S.W., Olson, R.J., Zettler, E.R., Goericke, R., Waterbury, J.B. and Welschmeyer, N. A., 1988. A novel free-living prochlorophyte abundant in the oceanic euphotic zone. Nature, 334(6180): 340-343.

Coleman, M., Sullivan, M., Martiny, A., Steglich, C., Barry, K., Delong. E. *et al*., 2006. Genomic islands and the ecology and evolution of *Prochlorococcus*. Science, 311(5768): 1768-1770.

Corpe, W.A., Matsuuchi, L. and Armbruster, B., 1976. Secretion of adhesive polymers and attachment of marine bacteria to surface, in: J. M Sharpley and A. M. Kaplan, Ed., Proceedings of the Third International Biodegradation Symposium, 433-442 pp, London

Cottrell, M.T. and Kirchman, D.L., 2000a. Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence in situ hybridization. Appl Environ Microbiol, 66(12): 5116-5122.

Cottrell, M.T. and Kirchman, D.L., 2000b. Natural assemblages of marine Proteobacteria and members of the Cytophaga-Flavobacteria cluster consuming low- and high-molecular-weight dissolved organic matter. Appl Environ Microbiol, 66(4): 1692-1697.

Cottrell, M.T., Yu, L. and Kirchman, D.L., 2005. Sequence and expression analyses of Cytophaga-like hydrolases in a Western Arctic metagenomic library and the Sargasso Sea. Appl Environ Microbiol, 71(12): 8506-8513.

Davis Jr., R A., 1972. Principles of Oceanography. Addison-Wesley Publishing Company, Philippines.

de la Torre, J R., Christianson, L.M., Béjà, O., Suzuki, M.T., Karl, D M., Heidelberg, J. *et al*., 2003. Proteorhodopsin genes are distributed among divergent marine bacterial taxa. Proc Natl Acad Sci U.S.A, 100(22): 12830-12835.

DeLong, E F., Franks, D.G. and Alldredge, A.L., 1993. Phylogenetic diversity of aggregate-attached vs. free-Living marine bacterial Assemblages. Limnol Oceanogr, 38(5): 924-934.

Dobrindt, U., Hochhut, B., Hentschel, U. and Hacker, J., 2004. Genomic islands in pathogenic and environmental microorganisms. Nat Rev Microbiol, 2(5): 414-424.

Doolittle, W.F., 1999. Lateral genomics. Trends Cell Biol, 9(12): M5-M8.

Ducklow, H.W., 1983. Production and fate of bacteria in the oceans. BioScience, 33(8): 494-501.

Dufresne, A., Garczarek, L. and Partensky, F., 2005. Accelerated evolution associated with genome reduction in a free-living prokaryote. Genome Biol, 6(2).

Engel, A., Thoms, S., Riebesell, U., Rochelle-Newall, E. and Zondervan, I., 2004. Polysaccharide aggregation as a potential sink of marine dissolved organic carbon. Nature, 428(6986): 929-932.

Falkowski, P., Scholes, R.J., Boyle, E., Canadell, J., Canfield, D., Elser, J. *et al*., 2000. The global carbon cycle: a test of our knowledge of earth as a system. Science, 290(5490): 291-296.

Falkowski, P G., Barber, R.T. and Smetacek, V., 1998. Biogeochemical controls and feedbacks on ocean primary production. Science, 281(5374): 200-206.

Falkowski, P.G., Fenchel, T. and Delong, E.F., 2008. The Microbial engines that drive Earth's biogeochemical cycles. Science, 320(5879): 1034-1039.

Fenchel, T., 1988. Marine plankton food chains. Ann Rev Ecol Syst, 19(1): 19-38.

References

Field, C.B., Behrenfeld, M.J., Randerson, J.T. and Falkowski, P., 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. Science, 281(5374): 237-240.

Finegold, S., Sutter, V. and Mathisen, G., 1983. Normal indigenous intestinal flora, in: Hentges D.J (ed.), Human intestinal microflora in health and disease. Academic Press, New York, 3-31 pp.

Finn, R.D. Tate, J., Mistry, J., Coggill, P. C., Sammut, S J., Hotz, H.-R. *et al*., 2008. The Pfam protein families database. Nucleic Acids Res, 36(suppl 1): D281-D288.

Fogg, G.E., 1958. Extracellular products of phytoplankton and the estimation of primary production. Rapp. P.-r. Réun. Cons. perm. int. Explor. Mer. , 144: 56-60

Frigaard, N.-U., Martínez, A., Mincer, T.J. and DeLong, E.F., 2006. Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. Nature, 439(7078): 847-850.

Fuhrman, J.A., Schwalbach, M.S. and Stingl, U., 2008. Proteorhodopsins: an array of physiological roles? Nat Rev Microbiol, 6(6): 488-494.

Garland, J. and Mills, A., 1991. Classification and characterization of heterotrophic microbial communities on the basis of patterns of community-level sole-carbon-source utilization. Appl Environ Microbiol, 57(8): 2351-2359.

Giovannoni, S. and Rappé, M., 2000. Evolution, diversity and molecular ecology of marine prokaryotes, in: Kirchman, D. L., (Ed.), Microbial Ecology of the Ocean. Wiley-Liss, New York, 47-84 pp.

Giovannoni, S.J., Foster, R.A., Rappé, M.S. and Epstein, S., 2007. New cultivation strategies bring more microbial plankton species into the laboratory. Oceanography, 22(2): 62-69.

Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D. *et al*., 2005. Genome streamlining in a cosmopolitan oceanic bacterium. Science, 309(5738): 1242-1245.

Glöckner, F.O., Fuchs, B. M. and Amann, R., 1999. Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence in situ hybridization. Appl Environ Microbiol, 65(8): 3721-3726.

Glöckner, F.O., Kube, M., Bauer, M., Teeling, H., Lombardot, T., Ludwig, W. *et al.*, 2003. Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. . Proc Natl Acad Sci U.S.A, 100(14): 8298-8303.

Gómez-Consarnau, L., González, J.M., Coll-Lladó, M., Gourdon, P., Pascher, T., Neutze, R. *et al*., 2007. Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. Nature, 445(7124): 210-213.

González, J.M. and Moran, M.A., 1997. Numerical dominance of a group of marine bacteria in the alpha-subclass of the class Proteobacteria in coastal seawater. Appl Environ

Microbiol, 63(11): 4237-4242.

González, J.M., Pedrós-Alió, C., and Gasol, J.M., 2008. Plancton bacteriano de los océanos. Investigación y Ciencia, December: 76-84.

Hacker, J., Bender, L., Ott, M., Wingender, J., Lund, B., Marre, R. *et al*., 1990. Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extra intestinal *Escherichia coli* isolates. Microb Pathog, 8(3): 213-225.

Hacker, J. and Kaper, J.B., 2000. Pathogenicity islands and the evolution of microbes. Annu Rev Microbiol, 54(1): 641-679.

Hedges, J.I., 1992. Global biogeochemical cycles: progress and problems. Mar Chem, 39(1-3): 67-93.

Hobbie, J.E., Daley, R.J. and Jasper, S., 1977. Use of nuclepore filters for counting bacteria by fluorescence microscopy. Appl Environ Microbiol, 33(5): 1225-1228.

Hollibaugh, J.T. and Azam, F., 1983. Microbial degradation of dissolved proteins in seawater. Limnol Oceanogr, 28(6): 1104-1116.

Hou, S., Saw, J.H., Lee, K.S., Freitas, T.A., Belisle, C., Kawarabayasi, Y. *et al*., 2004. Genome sequence of the deep-sea gamma-proteobacterium *Idiomarina loihiensis* reveals amino acid fermentation as a source of carbon and energy. Proc Natl Acad Sci U.S.A, 101(52): 18036-18041.

Huse, S.M., Dethlefsen, L., Huber, J. A. Welch, D.M., Relman, D. A. and Sogin, M.L.2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS Genet, 4(11): e1000255.

Insam, H., 1997. Substrate utilization tests in microbial ecology. A preface to the special issue of the Journal of Microbiological Methods. J Microbiol Methods, 30(1-2).

Juhas, M., van der Meer, J.R., Gaillard, M., Harding, R., Hood, D., Crook, D., 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol Rev, 33(2): 376-393.

Johannes, R., Ke Ying, W., Gerhard, J. H. and Edward, F. D., 1998. High phylogenetic diversity in a marine-snow-associated bacterial assemblage. Aquat Microb Ecol, 14(3): 261-269.

Kanehisa, M. and Goto, S., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res, 28(1): 27-30.

Kepkay, P.E., 1994. Particle aggregation and the biological reactiviy of colloids. Mar Ecol Prog Ser, 109: 293-304.

Kimura, H., Young, C.R., Martínez, A. and DeLong, E.F., 2011. Light-induced transcriptional responses associated with proteorhodopsin-enhanced growth in marine flavobacterium. ISME J, 5:1641-1651.

Kirchman, D.L., 2002. The ecology of Cytophaga-Flavobacteria in aquatic environments. FEMS Microbiol Ecol, 39(2): 91-100.

References

Kirchman, D.L., 2000. Microbial Ecology of the Oceans. Wiley-Liss, New York.

Kirchman, D.L., 2008. Microbial Ecology of the Oceans. John Wiley & Sons, Inc., Hoboken, New Jersey.

Kirchman, D.L., Moran, X. A. G. and Ducklow, H., 2009. Microbial growth in the polar oceans-role of temperature and potential impact of climate change. Nat Rev Microbiol, 7(6): 451-459.

Koonin, E.V., 2005. Orthologs, paralogs, and evolutionary genomics.. Annu Rev Genet, 39(1): 309-338.

Krogh, A., Brown, M., Mian, I.S., Sjölander, K. and Haussler, D., 1994. Hidden Markov Models in computational biology: applications to protein modeling. J Mol Biol, 235(5): 1501-1531.

Langille, M.G.I. and Brinkman, F. S. L., 2009. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. Bioinformatics, 25(5):664-665.

Langille, M.G.I ., Hsiao, W. W. L. and Brinkman, F. S. L., 2010. Detecting genomic islands using bioinformatics approaches. Nat Rev Microbiol, 8(5): 373-382.

Margulis, L. and Case E., 2006. The Germs of Life. Our ancestors were bacterial communities. Orion Magazine.

Martens, E.C., Koropatkin, N.M., Smith, T.J. and Gordon, J.I., 2009. Complex glycan catabolism by the human gut microbiota: The Bacteroidetes Sus-like paradigm. J Biol Chem, 284(37): 24673-24677.

Martínez, J., Smith, D.C., Steward, G. F. and Azam, F., 1996. Variability in ectohydrolytic enzyme activities of pelagic marine bacteria and its significance for substrate processing in the sea. Aquat Microb Ecol, 10: 223-230.

McCarren, J. and DeLong, E.F., 2007. Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla. Environ Microbiol, 9(4): 846-858.

Moran, M.A., Buchan, A., González, J. M., Heidelberg, J.F., Whitman, W. B., Kiene, R. P. *et al*., 2004. Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. Nature, 432(7019): 910-913.

Morris, R.M., Nunn, B.L., Frazar, C., Goodlett, D.R., Ting, Y.S., Rocap ,G. *et al*., 2010. Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. ISME J, 4(5): 673-685.

Nagata, T., 2000. Production mechanisms of dissolved organic matter, in: Kirchman, D. L., (Ed.), Microbial Ecology of the Ocean. Wiley-Liss, New York, 121-152 pp.

Nagata, T., 2008. Organic matter-bacteria interactions in seawater, in: Kirchman, D. L., (Ed.), Microbial Ecology of the Oceans. John Wiley & Sons, Inc., Hoboken, New Jersey, 207-242pp

Nikrad, MP., Cottrell, MT., Kirchman, DL. (2012). Abundance and single-cell activity of heterotrophic bacterial groups in the western Arctic Ocean in summer and winter. Appl Environ Microbiol, doi:10.1128/AEM.07130-11.

Navarre, W.W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S, J. *et al*., 2006. Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. Science, 313(5784): 236-238.

Nicol, G.W. and Schleper, C., 2006. Ammonia-oxidising Crenarchaeota: important players in the nitrogen cycle? Trends Microbiol, 14(5): 207-212.

Oesterhelt, D. and Stoeckenius, W., 1973. Functions of a new photoreceptor membrane. Proc Natl Acad Sci U.S.A, 70: 2853-2857.

Orr, J.C., Fabry, V.J., Aumont, O., Bopp, L., Doney, S.C., Feely, R.A. *et al*., 2005. Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. Nature, 437(7059): 681-686.

Overbeek, R., Disz, T. and Stevens, R., 2004. The SEED: a peer-to-peer environment of genome annotation. CACM-Bioinformatics, 47(11): 4746-51.

Pedrós-Alió, C., 2006. Genomics and marine ecology. Int Microbiol, 9: 191-197.

Pinhassi, J., Sala, M.M., Havskum, H., Peters, F., Guadayol, O., Malits, A. *et al*., 2004. Changes in bacterioplankton composition under different phytoplankton regimens. Appl Environ Microbiol, 70(11): 6753-6766.

Pomeroy, L.R., 1974. The ocean's food web, a changing paradigm. BioScience, 24(9): 499-504.

Pommier, T., Canback, B., Riemann, L., Bostrom, K.H., Simu, K., Lundberg, P., *et al* ., 2007. Global patterns of diversity and community structure in marine bacterioplankton. Mol Ecol, 16(4): 867-880.

Porter, K.G. and Feig, Y. S., 1980. The Use of DAPI for identifying and counting aquatic microflora. Limnol Oceanogr, 25(5): 943-948.

Redfield A. C., 1934. On the proportions of organic derivations in seawater and their relation to the composition of plankton, in: R.J. Daniel Ed., James Johnstone Memorial Volume. University Press of Liverpool, 177-192 pp.

Reeves, A.R., Wang, G.R. and Salyers, A.A., 1997. Characterization of four outer membrane proteins that play a role in utilization of starch *by Bacteroides thetaiotaomicron*. J Bacteriol, 179(3): 643-649.

References

Richter, M., Lombardot, T., Kostadinov, I., Kottmann, R., Duhaime, M., Peplies, J. *et al*., 2008. JCoast - A biologist-centric software tool for data mining and comparison of prokaryotic (meta)genomes. BMC Bioinformatics, 9: 177.

Riedel, T., Tomasch, J., Buchholz, I., Jacobs, J, Kollenberg, M., Gerdts, G. *et al*., 2010. Constitutive expression of the proteorhodopsin gene by a Flavobacterium strain representative of the proteorhodopsin-producing microbial community in the North Sea. Appl Environ Microbiol, 76(10): 3187-3197.

Riemann, L., Steward, G.F. and Azam, F., 2000. Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. Appl Environ Microbiol, 66(2): 578-587.

Robinson, C., 2008. Heterotrophic bacterial respiration in: Kirchman, D. L., (Ed.), Microbial Ecology of the Oceans. John Wiley & Sons, Inc., Hoboken, New Jersey, 299-334 pp.

Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A. *et al*., 2003. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. Nature, 424(6952): 1042-1047.

Rusch, D., Halpern, A., Sutton, G., Heidelberg, K., Williamson, S., Yooseph, S. *et al*., 2007. The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biol, 5(3): e77.

Schultz, J., Milpetz, F., Bork, P. and Ponting, C.P., 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U. S.A., 95(11): 5857-5864.

Sharp, J.H., 1973. Size Classes of Organic Carbon in Seawater. Limnol Oceanogr, 18(3): 441-447.

Sharp, J.H., 1977. Excretion of organic matter by marine phytoplankton: do healthy cells do it? Limnol Oceanogr, 22(3): 3881-399.

Sharrock, R., 2008. The phytochrome red/far-red photoreceptor superfamily. Gen Biol, 9(8): 230.

Shipman, J., Berleman, J. and Salyers, A., 2000. Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*. J Bacteriol., 182(19): 5365-5372.

Smith, D.C., Simon, M., Alldredge, A.L. and Azam, F., 1992. Intense hydrolytic enzyme activity on marine aggregates and implications for rapid particle dissolution. Nature, 359(6391): 139-142.

Stein, L., 2001. Genome annotation: from sequence to biology. Nat Rev Genet, 2(7): 493-503.

Stewart, H.B., 1966. In Deep Challenge. Van Nostrand Company, Inc. Princeton, New Jersey.

Stewart, R.H., 2008. Introduction to Physical Oceanography. Department of Oceanography, Texas A&M University.

Sutherland, I.W., 1985. Biosynthesis and composition of gram-negative bacterial extracellular and wall polysaccharides. Ann Rev Microbiol, 39(1): 243-270.

Suzuki N. and Kato K., 1953. Studies on suspended material marine snow in the sea. Part I. Bull. Fac. Fisheries Hokkoaido Univ: 132-135.

Tatusov, R.L., Koonin, E. V. and Lipman, D. J., 1997. A Genomic perspective on protein families. Science, 278(5338): 631-637.

Thurman, H.V., 1997. Introductory Oceanography. Prentice Hall College, New Jersey, USA.

Venter, C., Remington, K., Heidelberg, J., Halpern, A., Rusch, D., Eisen, J., *et al*., 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science, 304(5667): 66-74.

Verdugo, P., Alldredge, A. L., Azam, F., Kirchman, D. L., Passow, U., Santschi, P., H. *et al*., 2004. The oceanic gel phase: a bridge in the DOM-POM continuum. Mar Chem, 92(1-4): 67-85.

Vernikos, G.S. and Parkhill, J., 2008. Resolving the structural features of genomic islands: A machine learning approach. Genome Res, 18(2): 331-342.

Waterbury, J.B., Watson, S. W., Guillard, R.R.L. and Brand, L.E., 1979. Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. Nature, 277(5694): 293-294.

Weinbauer, M.G. and Peduzzi, P., 1995. Effect of virus-rich high molecular weight concentrates of seawater on the dynamics of dissolved amino acids and carbohydrates. Mar Ecol Prog Ser, 127: 245-253.

Whitman, W.B., Coleman, D.C. and Wiebe, W.J., 1998. Prokaryotes: The unseen majority. Proc Natl Acad Sci U.S.A, 95(12): 6578-6583.

Wilhelm, S.W. and Suttle, C.A., 1999. Viruses and nutrient cycles in the sea. BioScience, 49(10): 781-788.

Williams, K.P., 2002. Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. Nucleic Acids Res, 30(4): 866-875.

Williams, P.M. and Druffel, E.R.M., 1987. Radiocarbon in dissolved organic matter in the central North Pacific Ocean. Nature, 330(6145): 246-248.

Woese, C.R., Fos, G.E., Zablen, L., Uchida, T., Bonen, L., Pechman, K., *et al*., 1975. Conservation of primary structure in 16S ribosomal RNA. Nature, 254: 83-86.

Woyke, T,, Xie, G., Copeland, A., González, J.M., Han, C., Kiss, H. *et al*., 2009. Assembling the marine metagenome, one cell at a time. PLoS ONE, 4(4): e5299.

Xie, G., Bruce, D.C., Challacombe, J.F., Chertkov, O., Detter, J.C., Gilna, P. *et al*., 2007. Genome sequence of the cellulolytic gliding bacterium *Cytophaga hutchinsonii.* Appl Environl

Microbiol, 73(11): 3536-3546.

Xu, J., 2006. Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. Mol Ecol, 15(7): 1713-1731.

Zehr, J.P. and Capone, D.G., 1996. Problems and promises of assaying the genetic potential for nitrogen fixation in the marine environment. Microb Ecol, 32(3): 263-281.

Zimmerman R. and Meyer-Reil, L.A., 1974. A new method for fluorescence staining of bacterial populations on membrane filters. Kieler Meeresforsch, 30: 24-27.

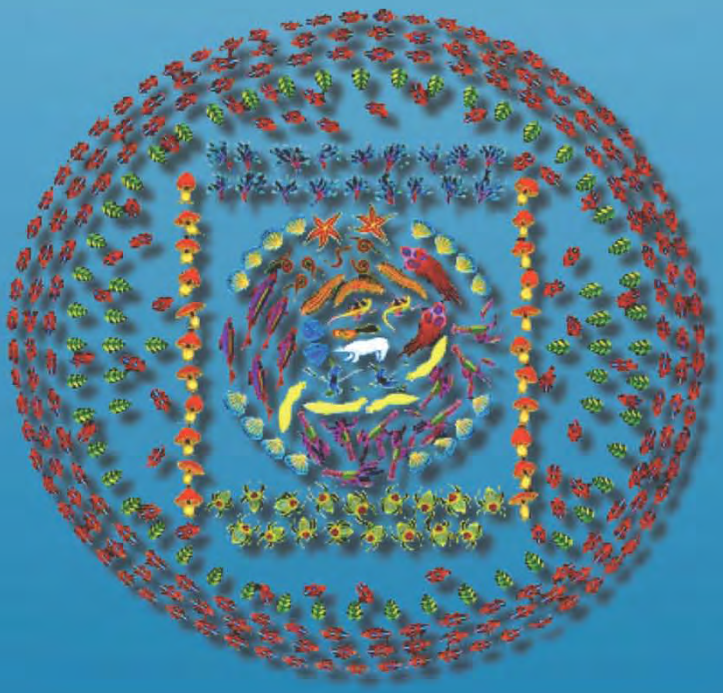
Yooseph, S., Nealson, K.H., Rusch, D.B., McCrow, J.P., Dupont, C.L., Kim, M. *et al.,*., 2010. Genomic and functional adaptation in surface ocean planktonic prokaryotes. Nature, 468(7320): 60-66.

ABBREVIATIONS INDEX

List of the abbreviations and acronyms used in this thesis:

AAnP: Aerobic Anoxygenic Photosynthesis
ABC transporters: ATP-binding cassette transporters
ATP: Adenosine triphosphate
BBMO: Blanes Bay Microbial Observatory
Bchl *a*: Bacteriochlorophyll *a*
BLAST: Basic Local Alignment Search Tool
CAZy: Carbohydrate-Active enzymes database
CFB: Cytophaga-Flavobacteria-Bacteroidetes
Chl *a*: Chlorophyll *a*
COG: Cluster of Orthologous Groups
CTD: Conductivity, temperature, depth
DAPI: 4,6-diamidino-2-phenylindole
DMSO: Dimethyl sulfoxide
DMSP: Dimethylsulfoniopropionate
DOC: Dissolved organic carbon
DOM: Dissolved organic matter
EPS: Exopolysaccharides
ExPASy: Expert Protein Analysis System
FISH: Fluorescence *in situ* hibridization
GH: Glycoside hydrolases
GI: Genomic Island
GO: Gene Ontology
GOS: Global Ocean Survey
HGT: Horizontal Gene Transfer
HMW-DOM: High molecular weight DOM
HP: Hypothetical protein
KEGG: Kyoto Encyclopedia of Genes and Genomes
NCBI: National Center for Biotechnology Information
OM: Organic matter
ORF: Open reading frame
PR: Proteorhodopsin
TOC: Total Organic Carbon

SERVERS, DATABASES AND INSTITUTIONS

NCBI (National Center for Biotechnology Information): http://www.ncbi.nlm.nih.gov
EMBL (The European Molecular Biology Laboratory): http://www.embl.de/
JGI (Joint Genome Institute): http://www.jgi.doe.gov/
KEGG (Kyoto Encyclopedia of Genes and Genomes): http://www.genome.jp/kegg/
ExPASy (Expert Protein Analysis System): http://expasy.org/
CAZy database (Carbohydrate-Active Enzymes): http://www.cazy.org/
MEROPS, the Peptidase Database: http://merops.sanger.ac.uk/
GenDB: http://www.cebitec.uni-bielefeld.de/groups/brf/software/gendb_info/
Silva: http://www.arb-silva.de/
BLAST (Basic Alignment Search Tool): http://blast.ncbi.nlm.nih.gov/Blast.cgi
JCoast (Comparative Analysis and Search Tool): http://www.jcoast.net/
IslandViewer: http://www.pathogenomics.sfu.ca/islandviewer/query.php
CAMERA database (Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis): http://camera.calit2.net/
Pfam (Protein family database): http://pfam.sanger.ac.uk/
Gen Ontology database: http://www.geneontology.org/
COG (Cluster of Orthologous Genes) database: http://www.ncbi.nlm.nih.gov/COG/
MAFFT: http://mafft.cbrc.jp/alignment/software/
RaxML Server: http://phylobench.vital-it.ch/raxml-bb/
iTOL: http://itol.embl.de/
gBlocks Server: http://molevol.cmima.csic.es/castresana/Gblocks.html
Betty and Moore Foundation: http://www.moore.org/
Craig Venter Institute: http://www.jcvi.org/
BBMO: http://www.icm.csic.es/bio/projects/icmicrobis/bbmo/

“Science will always be a search, never a real discovery.
It is a journey, never a finish”
Karl Popper