Marine Genomics 62 (2022) 100889

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Data

Complete genome of *Nocardioides aquaticus* KCTC 9944^T isolated from meromictic and hypersaline Ekho Lake, Antarctica

Kyuin Hwang [a], Hanna Choe [b], Kyung Mo Kim [a,*]

[a] Division of Polar Life Sciences, Korea Polar Research Institute, 26 Songdomirae-ro, Yeonsu-gu, Incheon 21990, Republic of Korea
[b] Biological Resource Center, Korea Research Institute of Bioscience and Biotechnology, 181 Ipsin-gil, Jeongeup 56212, Republic of Korea

ARTICLE INFO

Keywords:
Adaptation
Gradient
Salinity
Temperature
Tolerance

ABSTRACT

Nocardioides aquaticus KCTC 9944^T is an aerobic, non-motile, Gram-positive, psychrotolerant, non-spore-forming bacterium isolated from the surface water of Ekho Lake in the Vestfold Hills, East Antarctica. This meromictic lake separated from Antarctic seawater thousands of years ago exhibits steep gradients of salinity and temperature in the upper layer of the water column. The cells of *N. aquaticus* thriving in Ekho Lake are able to grow in wide ranges of temperature (3 to 43.5 °C) and salinity (0 to 15% NaCl). Here, we sequenced the complete genome of *N. aquaticus* KCTC 9944^T, aiming to better understand the adaptation of this bacterium to the strong environmental gradients at the molecular level. The genome consists of 4,580,814 bp (G + C content of 73.2%) with a single chromosome, 4432 protein-coding genes, 51 tRNAs and 2 rRNA operons. The genome possesses genes for the Entner-Doudoroff pathway, photoheterotrophy, the conversion of acetate to acetyl-CoA, gluconeogenesis, and energy storage that are all advantageous to oligotrophic bacteria. The presence of genes involved in osmotic balance, fatty acid desaturation, cold and heat shock responses, and the oxygen affinities of respiratory oxidases are likely associated with high tolerance to strong gradients of salinity, temperature and oxygen concentration.

1. Introduction

The Vestfold Hills are an ice-free area lying on the eastern edge of Prydz Bay on the Ingrid Christiansen Coast, East Antarctica (Zwartz et al., 1998). The contemporary landscape of this area formed approximately thousands of years ago by the retreat of the continental ice sheet followed by isostatic rebound lifted up faster than sea level rise (Gibson, 1999). The Vestfold Hills are diverse in topography, consisting of peninsulas, fjords and small hills (Labrenz et al., 1998), and they host ca. 300 lakes and ponds that vary in salinity (Rankin, 1998). Among these, Ekho Lake is a meromictic, hypersaline lake with 42 m deep (Labrenz et al., 1998; Lawson et al., 2000). The oxycline that separates top (mixolimnion) and bottom (monimolimnion) layers formed at approximately 24 m deep (Lawson et al., 2000). The following physical mechanisms result in the steep gradients of salinity and temperature with depth in the mixolimnion of this lake (Ferris et al., 1991). The surface water increases in temperature by the absorption of solar energy during summer and becomes more hypersaline by brine extrusion from the forming ice in winter. The warm, dense water beneath the ice cap sinks downwards, conveying salts and heat to the bottom of the mixolimnion. As a result, both salinity and temperature of the water column (50 g L^{-1} and 0 °C or so) increase with depth, and finally reach a maximum (165 g L^{-1} and 16–18 °C) in the vicinity of oxycline (Gibson, 1999).

A variety of bacterial groups including *Alpha-* and *Gammaproteobacteria*, *Cytophagales*, *Firmicutes*, *Cyanobacteria*, *Spirochaetales*, *Verrucomicrobia*, and *Actinobacteria* were discovered from Ekho Lake (Bowman et al., 2000; Labrenz and Hirsch, 2001). Among them, *Nocardioides aquaticus* EL-17K^T (= KCTC 9944^T = DSM 11439^T) that belongs to *Actinobacteria* was isolated from water samples at a depth of 1.0 m (temperature of 2.8 °C and salinity of 9.5% NaCl) of Ekho Lake on January 16, 1990 (Lawson et al., 2000). This bacterium is an aerobic, non-motile, Gram-positive, psychrotolerant, and non-spore-forming cocci or short rods (Lawson et al., 2000). A variety of carbon sources including acetate and D-glucose can be utilized, and nitrate can be assimilated under oxic conditions (Lawson et al., 2000). Cells grow at a temperature range of 3–43.5 °C (optimally 16–26 °C) and at a NaCl concentration range of 0–15% (optimally 1–6%) (Lawson et al., 2000), revealing that this bacterium has been adapted to the strong gradients of temperature and salinity. We here report the complete genome of *N. aquaticus* KCTC 9944^T, expecting that the genome will advance our

* Corresponding author.
E-mail address: kmkim@kopri.re.kr (K.M. Kim).

https://doi.org/10.1016/j.margen.2021.100889
Received 26 May 2021; Received in revised form 28 June 2021; Accepted 6 July 2021
Available online 13 July 2021

understanding of microbial adaptation to those environmental gradients at the molecular level.

2. Data description

General features of strain KCTC 9944^T were summarized in Table 1. Cells were grown for 10 days at 20 °C on Marine agar (Difco, USA). Genomic DNA was extracted using the i-genomic BYF mini kit (iNtRON Biotechnology, Republic of Korea). Genome sequencing of the strain was performed using PacBio RS II single-molecule real-time (SMRT) sequencing technology (Pacific Biosciences, USA). A 20-kb insert SMRTbell library was constructed and sequenced, yielding a 173-fold average genome coverage. De novo assembly of 69,770 subreads with 14,428 nucleotides on average (1,006,704,220 bp in total) was conducted using the hierarchical genome-assembly process (HGAP) of the SMRT Analysis v2.3.0 (Chin et al., 2013). Since a bacterial chromosome is mostly circular, both ends of a genome produced by the HGAP pipeline are overlapped in DNA sequence. This redundancy was manually removed to generate a unique stretch on both ends. Then, a new version for the genome was produced by bisecting the genome and switching the first half with the second. Raw PacBio reads were mapped to the newly generated genome using the resequencing module of the SMRT Analysis. This allows us to correct possible sequencing errors especially at both ends of the original version of the genome, where read mapping coverage is relatively lower.

Gene annotation was performed using Prokka v1.13 with *E-value* < 0.001 (Seemann, 2014). All predicted CDSs were BLAST-searched against COG and further analyzed using GhostKOALA (Kanehisa et al., 2016; Tatusov et al., 2000), with the aim to gain more insights into their molecular functions. Pseudogenes were determined by NCBI Microbial Genome Submission Check. Signal peptides and transmembrane helices were determined using SignalP v5 (Almagro Armenteros et al., 2019) and TMHMM v2.0 (Krogh et al., 2001), respectively. The circular genome map was constructed using CGView server v1.0 (Grant and Stothard, 2008).

The complete genome of *N. aquaticus* KCTC 9944^T with the average read concordance of 100% is composed of a circular chromosome of 4,580,814 bp with 73.2% G + C contents (Table 2). No plasmid was detected. In total, 4432 protein coding genes, 2 rRNA operons, 51 tRNAs and a tmRNA were found (Fig. 1; Table 2). 49 out of 4432 protein coding genes (1.10%) were annotated as pseudogenes. Signal peptides and transmembrane helices were found in 481 proteins (10.85%) and in 1057 proteins (23.85%), respectively.

Genes involved in oxidizing inorganic electron donors (e.g., reduced sulfurs, hydrogen, ammonia and nitrite) were not found in the genome. And, none of carbon fixation pathways were found. These results indicate that this bacterium is a chemoorganoheterotroph that oxidizes organic matter to produce cellular energy. The presence of a gene encoding beta-glucosidase (locus tag ENKNEFLB_04373) allows this bacterium to obtain glucose from polysaccharides in surrounding environments. The genomic potential of glucose utilization is also supported by the previous physiological characterization that the *N. aquaticus* EL-17K feeds on glucose (Lawson et al., 2000). A glucose-specific PTS system was absent in the genome. Instead, a transporter belonging to sodium/glucose transporter family (ENKNEFLB_04255) is thought to play a role in importing glucose into a cell. The genome exhibits the complete sets of genes for Embden-Meyerhof-Parnas (EMP), Entner-Doudoroff (ED), and pentose phosphate pathways. The EMP pathway (e.g., ENKNEFLB_01540) produces two ATP per glucose while the ED pathway (e.g., ENKNEFLB_00896) produces only one. On the other hand, ED requires less enzymatic protein than EMP, showing that there is a tradeoff between ATP yield and protein costs (Flamholz et al., 2013). This implies that ED and EMP pathways are suitable for oligotrophic and copiotrophic environments, respectively. In fact, *N. aquaticus* EL-17K grows well on oligotrophic PYGV (peptone, yeast extract, glucose and vitamins) agar (Lawson et al., 2000). Therefore, it is likely that this bacterium is able to switch between EMP and ED pathways depending on nutrient conditions.

The dull orange colony color of *N. aquaticus* EL-17K (Lawson et al., 2000) hints at the possibility that this bacterium absorbs sunlight via pigments and produces a proton motive force using rhodopsins when organic matter is not available for ATP generation (Gómez-Consarnau et al., 2019). As expected, a proteorhodopsin gene (ENKNEFLB_02146) was found in the genome. And, the neighboring gene (ENKNEFLB_02147) encodes a beta-carotene 15,15′-dioxygenase (EC 1.13.11.63 detected by GhostKOALA) that produces a chromophore retinal by cleaving beta-carotene. Therefore, *N. aquaticus* EL-17K may utilize the retinal-rhodopsin complex to generate ATP during starvation.

The genome revealed that acetate can be taken up via a cation/acetate symporter (ENKNEFLB_00406) and be converted into acetyl-CoA by an acetyl-CoA synthetase (ENKNEFLB_00225). Then, acetyl-CoA initiates the TCA cycle to produce ATP and NADH. It was reported

Table 1
General features of *Nocardioides aquaticus* KCTC 9944^T and MIGS mandatory information.

Item	Description
General features	
Classification	Domain *Bacteria* Phylum *Actinobacteria* Class *Actinobacteria* Order *Propionibacteriales* Family *Nocardioidaceae* Genus *Nocardioides*
Type strain	KCTC 9944^T
Gram strain	Positive
Cell shape	Cocci or Rods
Motility	Non-motile
Temperature range	3–43.5 °C, optimally at 16–26 °C
Salinity range	NaCl, 0 to 15%
pH range	5.5 to 9.5
MIGS data	
Submitted to INSDC	Accession number CP075371
Investigation type	Bacteria
Project name	Complete genome sequence of *Nocardioides aquaticus* KCTC 9944
Geographic location	Ekho Lake, Antarctica
Lat_lon	68.52 S, 78.27 E
Depth	1 m
Collection date	1990–01
Environment (biome)	Polar biome (ENVO:01000339)
Environment (feature)	Meromictic lake (ENVO:00000199)
Environment (material)	Hypersaline water (ENVO:00002012)
Environment (package)	Lake water (ENVO:04000007)
Isolation and growth conditions	PMID: 10930074
Sequencing	
Sequencing platform	PacBio RS II with P6-C4 chemistry
Fold coverage	173.09×
Assembler	SMRT Analysis v2.3.0
Annotation source	Prokka v1.13

Table 2
Genomic features of *N. aquaticus* KCTC 9944^T.

Attribute	Value
Genome features	
Genome size (bp)	4,580,814
Protein coding region (bp)	4,156,666
G + C content (%)	73.2
Chromosome	1
Total genes	4489
Protein coding genes	4432
rRNA operons	2 (16S–23S-5S)
tRNA	51
Pseudogenes	49
Genes assigned to COGs	3514
Genes with signal peptides	481
Genes with transmembrane helices	1057

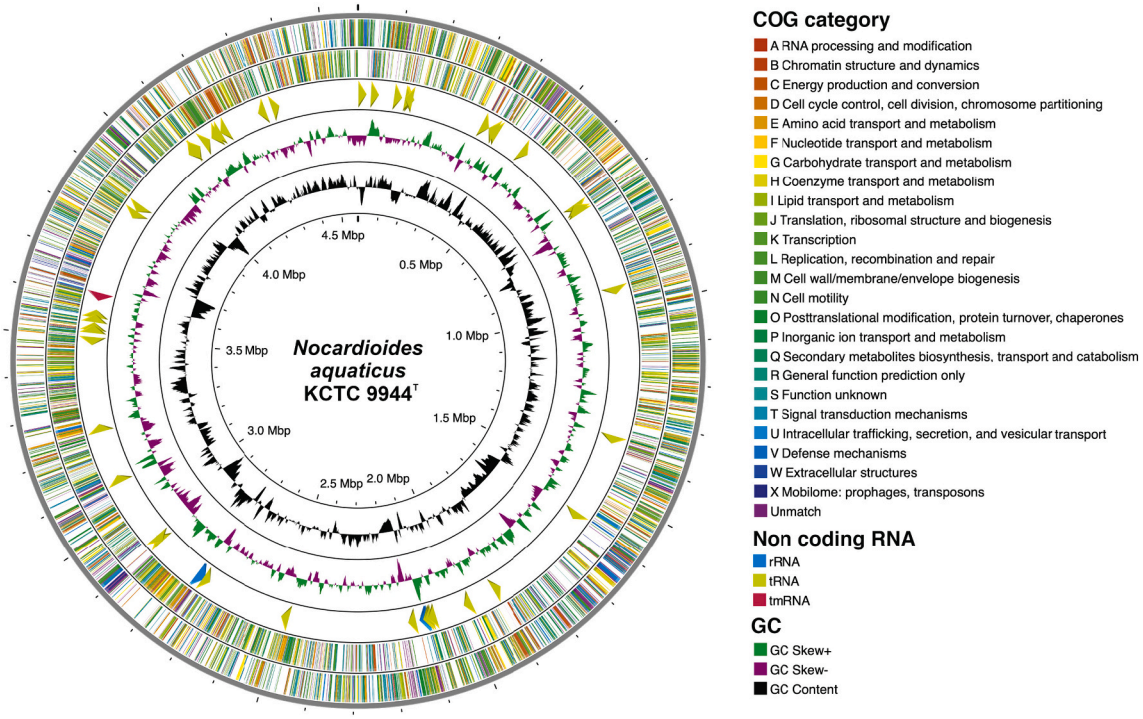

Fig. 1. Circular map of the *N. aquaticus* KCTC 9944^T genome. From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand, RNA genes (rRNAs blue, tRNA pale green, tmRNA red), GC skew (green/violet), and GC content (black). RNA genes on forward and reverse strands were labeled with clockwise and counterclockwise arrows, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that *N. aquaticus* EL-17K is able to utilize acetate for cell growth (Lawson et al., 2000). Therefore, both physiological and genomic data support the possibility that this strain can generate cellular energy by metabolizing acetate as an alternative carbon source when glucose is depleted in the habitat. Given that genes constituting the TCA cycle are completely present in the genome, the presence of genes for C4-dicarboxylic acid transporters (ENKNEFLB_00235 and _01451) supports the ability that this strain can metabolizes succinate and malate, as physiologically tested before (Lawson et al., 2000). All the genes for gluconeogenesis were found in this genome (e.g., ENKNEFLB_03713). Essential precursors of amino acids, nucleotides and lipids that should be produced by glycolysis are thought to be synthesized by gluconeogenesis, when cells grow on acetate in the absence of sugars.

The genome possesses genes for both cytochrome *c* oxidase (e.g., ENKNEFLB_01499) and cytochrome *bd* complex (ENKNEFLB_02185 and _02186) that exhibit low and high affinity for O_2, respectively (Gong et al., 2018). This allows this bacterium to respire in both oxygen-repleted and -depleted environments. Although this strain is known to assimilate nitrate under oxic conditions, the ability of nitrate dissimilation has not yet been reported (Lawson et al., 2000). In the genome, we found a complete set of genes involved in dissimilatory nitrate reduction to ammonia (DNRA; e.g., ENKNEFLB_00403), indicating that this bacterium is likely able to respire anaerobically using nitrate as a terminal electron acceptor. In addition, it seems that ammonia produced by DNRA diffuses out of cells of this strain and serves as an electron donor for microbial nitrification in Ekho Lake.

Among four major types of the Polyhydroxyalkanoate (PHA) synthase protein family, type II PHA synthases consist of two main subunits named PhaC and PhaE. Although *phaE* was not found in the genome, it is known that PhaC (ENKNEFLB_00936) is solely able to convert 3-hydroxybutyryl-CoA to polyhydroxybutyrate (PHB) and CoA (Müh et al., 1999), indicating that this strain possessing *phaC* has a potential to store excess carbon intracellularly at the nutrient-rich condition. However, glycogen storage seems to be absent since genes for glycogen synthase, NDP-glucose—starch glucosyltransferase, and starch synthase were not found in the genome. Owing to the presence of a polyphosphate kinase gene (ENKNEFLB_00731), this bacterium is likely capable of storing excessive energy as a form of polyphosphate (Kornberg et al., 1956).

Genes encoding high-affinity choline transporter, oxygen-dependent choline dehydrogenase, and betaine aldehyde dehydrogenase are organized in an operon-like manner (ENKNEFLB_03344 to _03346) in the genome. Among these enzymes, the two latter enzymes are responsible for betaine biosynthesis from choline. The existence of genes encoding trehalose import ATP-binding proteins, trehalose-binding lipoprotein, trehalose transport system permease proteins (ENKNEFLB_03716 to _03720), trehalose synthase/amylase (ENKNEFLB_01905) and betaine/ectoine transporter (ENKNEFLB_02175) implies that this bacterium is able to utilize the other major organic osmolytes (ectoine and trehalose) to reinforce persistence against steep changes of salinity in Ekho lake. In addition, transporters involved in sodium uptake (sodium/glucose cotransporter [ENKNEFLB_04255] and sodium-dependent dicarboxylate transporter [ENKNEFLB_02622]) and efflux (sodium, potassium, lithium, rubidium/proton antiporter [ENKNEFLB_04136]) may play a role in maintaining osmotic balance against varying salinity levels. Regarding thermal adaptation, we found four copies of genes encoding NADPH-dependent stearoyl-CoA Δ^9-desaturase (e.g., ENKNEFLB_01824). Since this enzyme is known to produce unsaturated fatty acids of the cell membrane (Chi et al., 2008), it would take part in regulation of membrane fluidity that is necessary for microbial survival against temperature changes. Additionally, it seems that cold and heat shock proteins (e.g., ENKNEFLB_00553 and _00717) reinforce the adaptation of this bacterium to a wide range of temperature changes.

The genome of *N. aquaticus* is equipped with a variety of genes involving high tolerance to strong gradients of temperature, salinity and oxygen concentration. Physiologically, this bacterium is known to be not motile (Lawson et al., 2000). Accordantly, the genome lacks genes for the biosynthesis of flagella and gas vesicle. These evidences imply that this bacterium must be planktonic and move along the water flow in the

mixolimnion of Ekho Lake. Genes present in this study is likely associated with the adaptation to dynamic changes of nutrient condition, temperature, salinity and oxygen concentration that are frequently encountered by the planktonic bacterium *N. aquaticus*.

3. Nucleotide sequence accession numbers

The complete genome sequence of *Nocardioides aquaticus* KCTC 9944^T has been deposited in GenBank under the accession number CP075371. This strain is available from the Korean Collection for Type Cultures (Jeongeup, Republic of Korea).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by Korea Polar Research Institute (grant number PE21130).

References

Almagro Armenteros, J.J., Tsirigos, K.D., Sonderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37, 420–423.

Bowman, J.P., McCammon, S.A., Rea, S.M., McMeekin, T.A., 2000. The microbial composition of three limnologically disparate hypersaline Antarctic lakes. FEMS Microbiol. Lett. 183, 81–88.

Chi, X., Yang, Q., Zhao, F., Qin, S., Yang, Y., Shen, J., Lin, H., 2008. Comparative analysis of fatty acid desaturases in cyanobacterial genomes. Comp. Funct. Genom. 2008.

Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., Turner, S.W., Korlach, J., 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods 10, 563–569.

Ferris, J.M., Gibson, J.A., Burton, H.R., 1991. Evidence of density currents with the potential to promote meromixis in ice-covered saline lakes. Palaeogeogr. Palaeoclimatol. Palaeoecol. 84, 99–107.

Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W., Milo, R., 2013. Glycolytic strategy as a tradeoff between energy yield and protein cost. Proc. Natl. Acad. Sci. 110, 10039–10044.

Gibson, J.A., 1999. The meromictic lakes and stratified marine basins of the Vestfold Hills, East Antarctica. Antarct. Sci. 11, 175–192.

Gómez-Consarnau, L., Raven, J.A., Levine, N.M., Cutter, L.S., Wang, D., Seegers, B., Arístegui, J., Fuhrman, J.A., Gasol, J.M., Sañudo-Wilhelmy, S.A., 2019. Microbial rhodopsins are major contributors to the solar energy captured in the sea. Sci. Adv. 5, eaaw8855.

Gong, X., Garcia-Robledo, E., Lund, M.B., Lehner, P., Borisov, S.M., Klimant, I., Revsbech, N.P., Schramm, A., 2018. Gene expression of terminal oxidases in two marine bacterial strains exposed to nanomolar oxygen concentrations. FEMS Microbiol. Ecol. 94, fiy072.

Grant, J.R., Stothard, P., 2008. The CGView server: a comparative genomics tool for circular genomes. Nucleic Acids Res. 36, W181–W184.

Kanehisa, M., Sato, Y., Morishima, K., 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. 428, 726–731.

Kornberg, A., Kornberg, S., Simms, E.S., 1956. Metaphosphate synthesis by an enzyme from Escherichia coli. Biochim. Biophys. Acta 20, 215–227.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580.

Labrenz, M., Hirsch, P., 2001. Physiological diversity and adaptations of aerobic heterotrophic bacteria from different depths of hypersaline, heliothermal and meromictic Ekho Lake (East Antarctica). Polar Biol. 24.

Labrenz, M., Collins, M.D., Lawson, P.A., TINDALL, B.J., BRAKER, G., HIRSCH, P., 1998. Antarctobacter heliothermus gen. nov., sp. nov., a budding bacterium from hypersaline and heliothermal Ekho Lake. Int. J. Syst. Evol. Microbiol. 48, 1363–1372.

Lawson, P.A., Collins, M.D., Schumann, P., Tindall, B., Hirsch, P., Labrenz, M., 2000. New LL-diaminopimelic acid-containing Actinomycetes from Hypersaline, Heliothermal and Meromictic Antarctic Ekho Lake: Nocardioides aquaticus sp. nov. and Friedmannielly lacustris sp. nov. Syst. Appl. Microbiol. 23, 219–229.

Müh, U., Sinskey, A.J., Kirby, D.P., Lane, W.S., Stubbe, J., 1999. PHA synthase from Chromatium vinosum: cysteine 149 is involved in covalent catalysis. Biochemistry 38, 826–837.

Rankin, L.M., 1998. The Ecology and Taxonomy of Synechococcus from Saltwater Lakes in the Vestfold Hills. University of Tasmania, Antarctica.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., Koonin, E.V., 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28, 33–36.

Zwartz, D., Bird, M., Stone, J., Lambeck, K., 1998. Holocene sea-level change and ice-sheet history in the Vestfold Hills, East Antarctica. Earth Planet. Sci. Lett. 155, 131–145.