

# SCIENTIFIC REPORTS



OPEN

## Genome-guided Investigation of Antibiotic Substances produced by *Allosalinactinospora lopnorensis* CA15-2<sup>T</sup> from Lop Nor region, China

Received: 10 May 2015  
Accepted: 14 December 2015  
Published: 11 February 2016

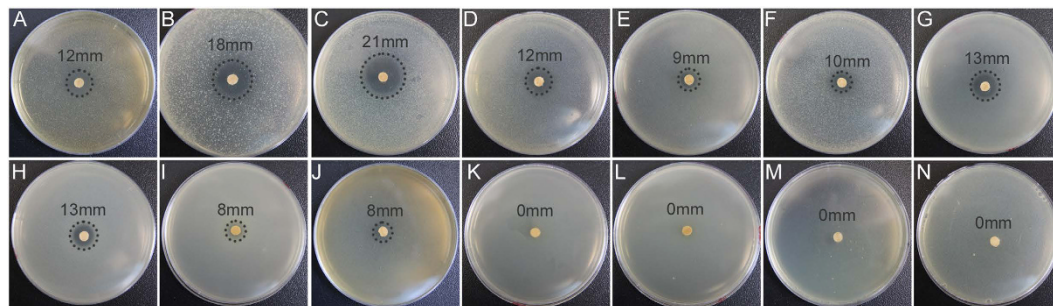
Chen Huang<sup>1,\*</sup>, Ross Ka-Kit Leung<sup>2,5,\*</sup>, Min Guo<sup>1,\*</sup>, Li Tuo<sup>3</sup>, Lin Guo<sup>3</sup>, Wing Wai Yew<sup>2</sup>,  
Inchio Lou<sup>4</sup>, Simon Ming Yuen Lee<sup>1</sup> & Chenghang Sun<sup>3</sup>

Microbial secondary metabolites are valuable resources for novel drug discovery. In particular, actinomycetes expressed a range of antibiotics against a spectrum of bacteria. In genus level, strain *Allosalinactinospora lopnorensis* CA15-2<sup>T</sup> is the first new actinomycete isolated from the Lop Nor region, China. Antimicrobial assays revealed that the strain could inhibit the growth of certain types of bacteria, including *Acinetobacter baumannii* and *Staphylococcus aureus*, highlighting its clinical significance. Here we report the 5,894,259 base pairs genome of the strain, containing 5,662 predicted genes, and 832 of them cannot be detected by sequence similarity-based methods, suggesting the new species may carry a novel gene pool. Furthermore, our genome-mining investigation reveals that *A. lopnorensis* CA15-2<sup>T</sup> contains 17 gene clusters coding for known or novel secondary metabolites. Meanwhile, at least six secondary metabolites were disclosed from ethyl acetate (EA) extract of the fermentation broth of the strain by high-resolution UPLC-MS. Compared with reported clusters of other species, many new genes were found in clusters, and the physical chromosomal location and order of genes in the clusters are distinct. This study presents evidence in support of *A. lopnorensis* CA15-2<sup>T</sup> as a potent natural products source for drug discovery.

Natural products represent a leading source for drug candidates, especially in the anti-bacterial field<sup>1,2</sup>. Actinomycetes, regarded as a monophyletic branch of bacteria, are fungi-like bacteria forming long filaments; most importantly, these microbes tend to produce secondary metabolites such as polyketides and non-ribosomal peptides<sup>3,4</sup>. Many of these compounds have been successfully isolated and transformed into useful drugs, antibiotics, or chemotherapeutic agents<sup>5,6</sup>. Accordingly, there are more than 22,000 known microbial secondary metabolites, 70% of which are produced by actinomycetes, 20% from fungi, 7% from *Bacillus spp.* and 1–2% by other bacteria<sup>7</sup>.

There has been a large push towards research on the anti-bacterial mechanisms of these microbes, which can be traced to the discovery of streptomycin. Presently, researchers have been especially keen to explore new drugs by improving or activating the metabolites of microorganisms. Olano, *et al.*<sup>8</sup> applied genetic approaches to bioactive secondary metabolites producing actinomycetes to improve the activity of relevant enzymes. Ito, *T*<sup>8</sup> modified this method to improve efficacy and reduce the toxicity of actinomycete metabolites by creating new analogs using mutasynthesis. In addition, Yassin and Mankin found 77 new targets after random mutation in the RNA of a large ribosomal subunit<sup>9</sup>. Another difficulty for experimental work to resolve is that culture conditions can influence the expression of gene clusters in microbes.

<sup>1</sup>State Key Laboratory of Quality Research in Chinese Medicine and Institute of Chinese Medical Sciences, University of Macau, Macao, China. <sup>2</sup>Stanley Ho Centre for Emerging Infectious Diseases, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China. <sup>3</sup>Department of Microbial Chemistry, Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China. <sup>4</sup>Faculty of Science and Technology, Department of Civil and Environmental Engineering, University of Macau, Macao, China. <sup>5</sup>School of Public Health, The University of Hong Kong, Hong Kong. <sup>\*</sup>These authors contributed equally to this work. Correspondence and requests for materials should be addressed to S.M.Y.L. (email: simonlee@umac.mo) or C.S. (email: sunchenghang@imb.pumc.edu.cn or chenghangsun@hotmail.com)



**Figure 1. Comparison of inhibition zone diameters produced by disks among fourteen bacteria or fungi.** (A) *Candida albicans* CCTCC AY93025; (B) *Acinetobacter baumannii* 2799; (C) *Acinetobacter baumannii* ATCC 19606; (D) *Escherichia coli* ATCC 25922; (E) *Escherichia coli* ATCC 2800; (F) *Klebsiella pneumoniae* ATCC 700603; (G) *Klebsiella pneumoniae* ATCC 10031; (H) *Staphylococcus aureus* ATCC 25923; (I) *Staphylococcus aureus* 2641; (J) *Cryptococcus laurentii* CCTCC AY91013; (K) *Enterococcus faecalis* ATCC 29212; (L) *Enterococcus faecalis* ATCC 33186; (M) *Pseudomonas aeruginosa* 2774; (N) *Pseudomonas aeruginosa* ATCC 27853.

At present, most genes and pathways involved in synthesizing various antibiotics have been identified, for example, the DOBISCUIT database collects around one hundred diverse antibiotics biosynthetic gene clusters. Moreover, Blin, *et al.* summarized previous findings on gene clusters related to different kinds of antibiotics and secondary metabolites, and automated the identification of the gene clusters encoding secondary metabolite production<sup>10</sup>. These two resources and alike facilitate comparison of potential antibiotics-producing gene clusters. However, “Me-Too” drugs may not help much<sup>11</sup>. The unremitting emergence of antibiotics-resistance also draws attention to new antibiotics source from exotic origins, such as mangrove<sup>12</sup>, and marine salterns<sup>7</sup>. Therefore, microbes from those environments may have great potential for novel antibiotics discovery<sup>7,13–15</sup>. It is particularly notable that microorganisms existing under extreme environmental conditions should exhibit distinctive features and abilities, and thus should be the most likely to yield unique and powerful bioactive compounds.

Schroeckh *et al.* found that *Aspergillus nidulans*, when co-cultured with bacteria, could produce several novel metabolites<sup>16</sup>. The capacity of secondary metabolic biosynthesis of microorganisms is always greater than that observed by fermentation<sup>4</sup>. Microbial genome analysis is an effective way of novel natural product producing gene discovery, due to clustering of biosynthetic genes of secondary metabolites e.g. polyketides, in genome. Beyond the known gene clusters, exploring the properties of the numerous orphan genes is regarded as the biggest challenge to finding new bioactive metabolites<sup>17</sup>. After genome sequencing, Udway *et al.* found that approximately 9.9% of the marine actinomycete *Salinispora tropica* genome is related to natural product assembly, and 17 novel biosynthetic loci<sup>4</sup>. Hu *et al.* characterized two new cryptic sesquiterpene synthases in the *Streptomyces clavuligerus* genome by combining genetic and biochemical techniques<sup>18</sup> while Soror *et al.* characterized an unusual member of the hormone-sensitive lipase family of esterases in *Streptomyces coelicolor* A3 (2)<sup>19</sup>. Lautru *et al.* found that in *Streptomyces coelicolor* M145 several gene clusters encoding new non-ribosomal peptide synthetase (NRPS) systems are not associated with known secondary metabolites; furthermore, they isolated and determined the structure of a new tris-hydroxamate tetrapeptide iron chelator coelichelin using a genome mining approach and guided by substrate predictions<sup>20</sup>.

To the best of our knowledge, strain *Allosalinactinospora lopnorenensis* CA15-2<sup>T</sup> is the first new actinomycete in genus level found in the Lop Nor region of the Xinjiang province of China, which is famous for its high temperature, salinity and drought<sup>21</sup>. Antimicrobial assays revealed that the strain could inhibit the growth of certain types of bacteria, including *Acinetobacter baumannii* and *Staphylococcus aureus*, thus, using a genomics approach to unveil and explore the potential abilities of the type strain of the new species and new genus shall guide subsequent experiments for the discovery of novel natural bioactive compounds for helping human tackle bacterial infections.

## Results

**Culture of strain CA15-2<sup>T</sup> and its Antimicrobial Assay.** Better growth of strain *A.lopnorenensis* CA15-2<sup>T</sup> on R2A with 5% NaCl than without NaCl were observed at 28 °C pH 7.5 for 10 days (Figure S1) Antimicrobial assay shows that concentrated sample of ethyl acetate extract from the fermentation broth of strain *A. lopnorenensis* CA15-2<sup>T</sup> exhibits different degrees of inhibitory activity against two fungi and eight bacteria, in particular *Acinetobacter baumannii* 2799 and *Acinetobacter baumannii* ATCC19606 (Fig. 1).

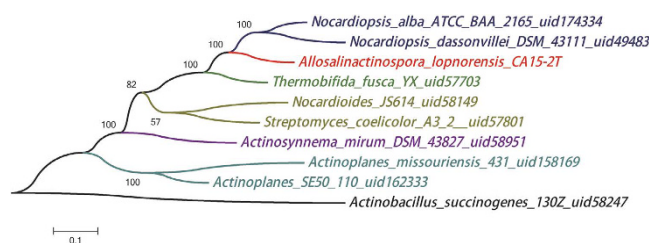
**Genome assembly and annotation.** Deep sequencing based on enzymatic digestion DNA fragmentation, yielded 1,945,282 raw reads (average read length = 137 bp). After removing short, low quality and suspected-plasmid reads, 260 contigs of more than 5.8 Mb were obtained. The average coverage was 45.1 fold and the G + C content was approximately 69.61%, consistent with the result (69.60%) obtained by reverse-phase HPLC<sup>22</sup>. Another round of sequencing with DNA fragmentation by sonication generated 3,317,091 raw reads (average read length = 222 bp), which were subsequently assembled into 1,456 contigs. In order to obtain a better assembly result, we combined the sequencing reads of the two different DNA fragmentation methods, resulting in a total of 5,200,564 reads, or 1,002,084,924 base-pairs (average read length 179.43 bp). The combined reads

Sequence data	Enzymatic digestion			Sonication		
	Insert size (bp)	High-quality data (Mb)	Sequence depth (fold coverage)	Insert size (bp)	High-quality data (Mb)	Sequence depth (fold coverage)
Raw data	5–353	266.2	45.13	5–395	736.3	124.83
Filter data	50–353	202.9	34.45	50–395	684.9	116.28
Data merge						
Assembly statistics	ContigN50 (bp)	61350				
	Max length (bp)	231302				
	GC content (%)	69.61				
	Total size (Mb)	5.89				

**Table 1. Genome sequencing and assembly statistics.**

Contigs	233
DNA, total number bases	5894259
Coding density (%)	85.4
G + C content (%)	69.61
Genes, total number	5662
Protein coding genes	5605
Total RNA genes	57
5S rRNA	1
16S rRNA	1
23S rRNA	1
tRNA	54

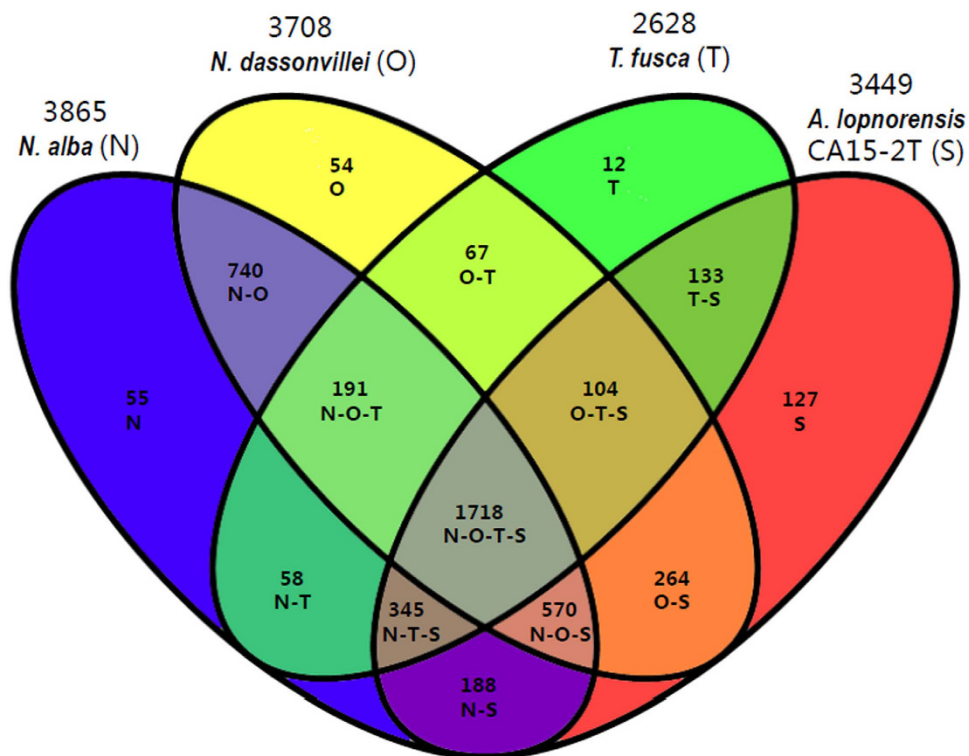
**Table 2. General features of the genome of *A. lopnorenensis* CA15-2<sup>T</sup>.**



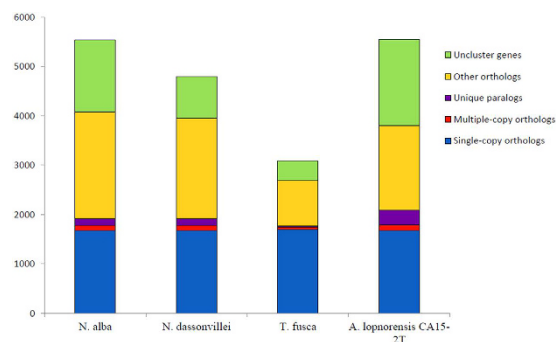
**Figure 2. Comparison of ten housekeeping genes from strain *A. lopnorenensis* CA15-2<sup>T</sup> with other orthologous.** Phylogenetic tree based on maximum likelihood method of a multi-alignment of ten housekeeping genes, respectively. The number of bootstrap replication was set to 500.

were assembled into 233 contigs of 5,897,123 bp, and an average G + C content of 69.61% (Table 1). The raw data and the total shotgun assembly (TSA) were submitted and archived in the GenBank under the accession number LAJ00000000 and SRS881470, under the BioProject nr PRJNA278354 and biosample nr SAMN03418058. Detailed information on sequencing and assembly is shown in Table 1. Specifically, more than 5,000 ORFs were predicted, with 5,549 protein-coding genes subjected to further annotation analysis (Figure S2). A total of 4,717 putative protein-coding genes had homologs identified in the *nr* database, with 2,987 sequences assigned to 22 functional categories by egg NOG classifications (Figure S3, A–V). The majority of these protein sequences were found to involve, in general function terms, energy and transcription. Fifty-seven RNA coding sequences were detected, including 5S, 16S and 23S rRNAs and the remaining consists of tRNAs (Table 2). The functions of remaining 832 genes cannot be detected by similarity-based search methods, suggesting that *A. lopnorenensis* CA15-2<sup>T</sup> may carry a novel gene pool.

**Phylogenetic and comparative analysis.** Phylogenetic results based on 16S rRNA showed that *A. lopnorenensis* CA15-2<sup>T</sup>, together with *Nocardioopsis alba* ATCC BAA2165, *Nocardioopsis dassonvillei* DSM43111, and *Thermobifida fusca* YX, are located at the same branch of the tree (Figure S4), which suggests that these species should have a close phylogenetic relationship to other species on the tree. The phylogeny constructed by house-keeping genes is in accordance with the 16S rRNA gene sequences (Fig. 2). Orthology analysis showed that 4,378 ortholog families are shared between at least two of the genomes from *N. alba* ATCCBAA2165, *N. dassonvillei* DSM43111, *T. fusca* YX or *A. lopnorenensis* CA15-2<sup>T</sup>; 1,718 (27.6%) orthologs were shared by all four species (Fig. 3). Moreover, there were 570 (13%) families shared among all of the *Nocardioopsis* species, 127 (2.7%)



**Figure 3.** Venn diagram displaying the distribution of shared gene families among *N. alba*, *N. dassonvillei*, *T. fusca* and *A. lopnorenسيس CA15-2<sup>T</sup>*. Numbers indicate the number of gene families in each cluster. The Venn diagram was created with R software.



**Figure 4.** Cumulative histogram displaying the proportion of different types orthologs genes among *N. alba*, *N. dassonvillei*, *T. fusca* and *A. lopnorenسيس CA15-2<sup>T</sup>*.

families that only belonged to the strain *A. lopnorenسيس CA15-2<sup>T</sup>*. For those genes involved in the orthologous family, all four species had similar numbers (approximately 1,600) of single-copy orthologous genes (Fig. 4). In addition, 311 unique paralogous genes, belonging to 127 unique families, were found in *A. lopnorenسيس CA15-2<sup>T</sup>*, by contrast, *N. dassonvillei* has 128 unique paralogous genes, *N. alba* has 138 ones and *T. fusca* has only 26 unique paralogous genes. The larger amount of unique paralogous *A. lopnorenسيس CA15-2<sup>T</sup>* might associated with its unequalled features of adaptation to extreme environments and capacity of potential novel secondary metabolites biosynthesis.

Four genes encoding the enzymes for ectoine (1,4,5,6,tetra-2-methyl-4-pyrimidonecarboxylic acid) synthesis were identified. The four putative ectoine biosynthetic genes are located very close one after the other (Table 3) and the order of these four genes is highly similar to those of other halophiles of actinomycete, e.g. *Streptomyces coelicolor* A3 (2), *Nocardia farcinica* IFM10152 and *T. fusca* YX<sup>23</sup>, which provides evidence for the molecular basis of adaptation to a saline environment (Figure S1).

**Antibiotics-producing Capability Assessment.** Our antimicrobial assay revealed that *A. lopnorenسيس CA15-2<sup>T</sup>* inhibit the growth of specific bacteria, and gene cluster prediction also indicated that it may contain novel gene clusters for antibiotics synthesis. Here, a comparative analysis of *A. lopnorenسيس CA15-2<sup>T</sup>*

Gene ID	Coding enzyme	Residue length	Start position	End position
Ect-A	DABA-N $\gamma$ -acetyltransferase	160	38792	39271
Ect-B	DABA transaminase	422	37423	38688
Ect-C	Ectoine synthase catalyses	139	36904	37320
EcT-D	Ectoine hydroxylase	304	35829	36740

**Table 3. Detailed information of four genes encoding enzymes involved in Ectoine synthesis of *A. lopnorenensis* CA15-2<sup>T</sup> located in assembled sequence (contig00008).**

with 10 diverse antibiotic-producing bacteria, as well as 10 non-antibiotic-producing bacteria was conducted to assess the capacity of producing antibiotics, and a new measurement Q was proposed, in which not only matched similarity (identity value) and length (alignment length) were taken into account, but also the gene amount was used for normalization, in terms of the higher the gene number, the greater the probability of gene sets of this species matching the sequences of gene clusters involved in antibiotic production. The results indicated that the majority of Q values of the antibiotic-producing group were markedly larger than those of non-antibiotics-producing group (Tables 4 and 5). It should be noted that some gene clusters of biosynthetic antibiotics (e.g. Avermectin, Myxothiazol) showed no obvious difference between the antibiotics-producing and non-antibiotic-producing groups merely based on a count of the hit number of gene/PKS genes during the blast search (Table S1,S2), whereas exhibited obvious differences in Q values (Tables 4 and 5). Furthermore, statistical test repeated measures ANOVA for the comparison of antibiotics-producing and non-antibiotic-producing groups showed that the difference in Q value between the antibiotics-producing and non-antibiotic-producing groups is statistically significant. Concretely speaking, multivariate test indicated that the selection of gene cluster did not affect the efficiency of Q-value (P value = 0.001, Table S3), and test of Between-Subjects Effects showed that Q value between the antibiotics-producing and non-antibiotic-producing groups is significantly different (P value  $\leq 0.05$ , Table S4). On the other hand, twenty one bacteria including ten antibiotics-producing bacteria and ten non-antibiotic-producing bacteria as well as *A. lopnorenensis* CA15-2<sup>T</sup> was classified using principal component analysis (PCA) on Q value, which illustrated that *A. lopnorenensis* CA15-2<sup>T</sup> tend to cluster with antibiotics-producing group (Figure S5). Similarly, cluster and phylogenetic analysis of Q value revealed that the strain *A. lopnorenensis* CA15-2<sup>T</sup>, and the antibiotics-producing species, were placed in the same group (Figure S6,S7), pointing toward the potential presence of the antibiotics-producing capacity in the strain *A. lopnorenensis* CA15-2<sup>T</sup>.

**Detection of gene clusters involved in Antibiotics & Secondary Metabolites.** A total of 17 secondary metabolic biosynthesis gene clusters were identified in *A. lopnorenensis* CA15-2<sup>T</sup>, which are predicted for polyketide, non-ribosomal peptide, siderophore, terpenoid and other products (Table 6). The combined length of these gene clusters was estimated at 412 kb. Specifically, the analysis shows that the genome of *A. lopnorenensis* CA15-2<sup>T</sup> may contain at least six kinds of type I PKS gene clusters (Table 6), in which three core structures were identified using anti SMASH prediction. In the present study, the analysis revealed that four of the above PKS I gene clusters are likely to be involved in the biosynthesis of simocyclinone, apoptolidin, nigericin and oxazolomycin. The simocyclinone biosynthesis gene cluster of *A. lopnorenensis* CA15-2<sup>T</sup> contains 35 ORFs (Fig. 5), five of which encode modular PKS. These modular PKSs contain 13 catalytic domains in which acyl-chain elongation involves acyl carrier protein (ACP),  $\beta$ -ketoacyl-ACP reductase (KR),  $\beta$ -ketoacyl-ACP synthase (KS), acyltransferase (AT) and dehydratase (DH). In addition, the simocyclinone cluster of *A. lopnorenensis* CA15-2<sup>T</sup> is relatively similar to the existing simocyclinone cluster of *Nocardioopsis dassonvillei* subsp. *dassonvillei* DSM 43111 in aspects of amino acid sequence (Figure S8,S9), ORF order as well as composition, particularly with respect to several homologous putative proteins (white-striped arrows in Fig. 5) that may play a role in the synthesis of the simocyclinone-derived primer unit, which indicates that these clusters may be derived from donor microorganisms via horizontal gene transfer. Another type I PKS gene cluster may produce nigericin-like compound with high similarity to *Ktedonobacter racemifer* DSM 44963. The type I PKS cluster contains 19 ORFs, of which 5 encode modular PKS. Likewise, both clusters are composed of numerous hypothetical unknown function proteins (Fig. 6), and the majority of the key PKS genes share high amino acid similarity (Figure S10). Furthermore, an apoptolidin-like PKS I biosynthetic gene cluster was found in the *A. lopnorenensis* CA15-2<sup>T</sup> genome. Comparison of both clusters indicates that apoptolidin-like cluster has several novel members that are not present in the *Streptomyces bingchenggensis* BCW-1 cluster (Fig. 7), including oxidoreductase, short-chain dehydrogenase and phosphopantetheine-binding domain-containing protein. However, the apoptolidin cluster of *Streptomyces bingchenggensis* BCW-1 has more polyketide synthase than the *A. lopnorenensis* CA15-2<sup>T</sup> cluster. In addition to the abundance of the diverse PKS cluster, *A. lopnorenensis* CA15-2<sup>T</sup> harbors a non-ribosomal peptide synthetase (NRPS) pathway. The NRPS cluster is composed of 36 ORFs, of which 8 encode key PKS modular with 13 domains. Compared with the most-similar known cluster of *Streptomyces* sp. W007 contig 00127, the cluster of *A. lopnorenensis* CA15-2<sup>T</sup> has several unique ORFs including Hydrolase, Ketoacyl synthase and transcriptional regulators, which indicates that *A. lopnorenensis* CA15-2<sup>T</sup> may have the potential of synthesizing an oxazolomycin-analog (Fig. 8).

**Secondary metabolites Identified Directly by Mass Spectrometric Analysis.** Mass spectra of the ethyl acetate (EA) extract of the fermentation broth of the strain CA15-2<sup>T</sup> generally exhibited a large number of discrete ions ranging from m/z 50 to 1200 both in positive and in negative ionisation modes. From UPLC-MS

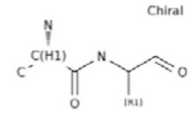
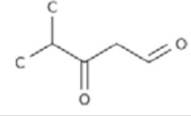
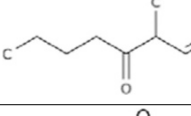
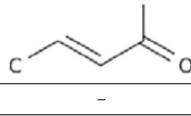
<i>species</i> Type	<i>Actino. mirum</i>	CA15-2 <sup>T</sup>	<i>Micro. aurantiaca</i>	<i>Nostoc. punctiforme</i>	<i>Pseudo. fluorescens</i>	<i>Sal. tropica</i>	<i>Sora. cellulosum</i>	<i>Strep. albus</i>	<i>Strep. avermitilis</i>	<i>Strep. coelicolor</i>	<i>Strep. collinus</i>
Aclacinomycin A	0.886	0.954	2.394	0.663	0.852	2.415	0.421	1.276	1.272	1.347	1.688
Amphotericin B	8.318	4.973	9.814	4.257	2.574	10.403	1.937	14.505	8.982	5.456	4.137
Ascomycin	4.658	2.972	4.662	2.000	1.898	6.622	1.259	5.280	3.966	3.300	2.925
Avermectin I	9.018	7.975	9.571	4.905	3.769	12.353	2.001	10.020	34.503	6.901	6.051
Chromomycin	1.284	1.175	2.154	1.121	0.876	2.404	0.680	1.711	1.342	1.228	1.908
Concanamycin A	5.701	4.223	7.396	2.987	2.130	7.773	1.468	7.469	6.334	3.923	3.973
Doxorubicin	0.974	1.239	2.447	1.081	1.022	2.805	0.689	1.330	1.533	1.586	1.913
Enterocin	1.204	1.723	1.766	0.932	1.119	2.540	0.739	1.487	1.356	1.404	1.621
Epothilone	3.409	2.635	4.345	3.698	2.143	4.207	1.811	4.118	3.440	2.538	2.829
Herbimycin A	6.029	2.926	4.820	1.719	1.612	4.452	1.143	4.302	3.318	2.692	2.728
Lasalocid	5.740	4.664	6.460	3.080	2.376	8.037	1.614	6.606	5.657	4.609	4.224
Mithramycin	1.508	1.486	2.138	1.152	0.993	2.526	0.757	1.965	1.516	1.396	2.176
Monensin	4.215	4.041	5.001	2.664	2.155	6.346	1.326	5.488	4.454	3.527	3.301
Mycolactone	0.910	0.607	1.736	0.502	0.301	1.842	0.206	1.140	1.389	0.690	0.472
Myxothiazol	5.914	5.202	7.503	7.772	5.389	8.406	4.345	6.442	5.162	5.000	6.028

**Table 4.** Q values of ten antibiotic-producing bacteria as well as *A. lopnorenensis* CA15-2<sup>T</sup> based on the blastp alignment against 15 gene clusters involved in antibiotic synthesis.

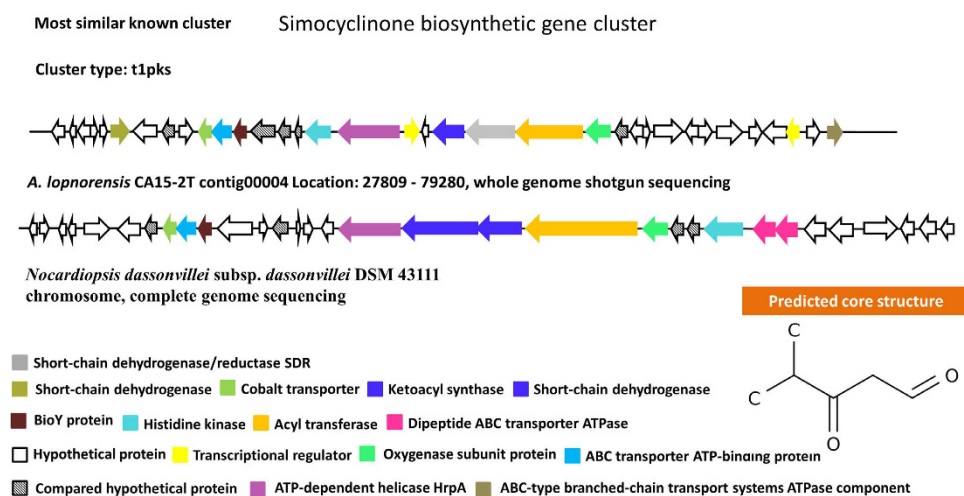
<i>Species</i> Type	<i>Bacillus. cereus</i>	<i>Burk. xenovorans</i>	<i>Entero</i>	<i>Kleb. pneumoniae</i>	<i>Magne. magneticum</i>	<i>Rhodop. palustris</i>	<i>Salmon. enterica</i>	<i>Shigella. dysenteriae</i>	<i>Xantho. axonopodis</i>	<i>Xantho. oryzae</i>
Aclacinomycin A	0.723	0.942	0.917	0.700	0.933	0.719	0.974	1.031	0.900	0.755
Amphotericin B	1.503	1.657	1.693	1.339	3.458	2.864	1.519	1.475	1.842	1.471
Ascomycin	1.077	1.190	0.796	0.779	1.783	1.819	0.973	0.895	1.496	0.743
Avermectin	1.244	1.294	1.377	1.232	3.903	3.345	1.270	1.418	1.392	1.166
Chromomycin	1.016	1.314	1.158	0.989	1.233	1.097	1.398	1.350	1.131	0.964
Concanamycin A	0.855	1.118	0.978	0.883	2.968	1.994	0.886	1.009	1.204	0.896
Doxorubicin	1.161	1.392	1.553	1.235	1.410	1.235	1.281	1.351	1.389	1.190
Enterocin	1.307	1.684	1.512	1.293	1.225	1.175	1.453	1.228	1.515	1.143
Epothilone	0.944	0.835	1.138	0.882	2.720	2.386	0.943	1.073	1.477	0.974
Herbimycin A	1.175	1.195	1.205	1.144	1.721	1.663	1.077	1.109	1.190	0.638
Lasalocid	1.027	1.043	1.178	0.962	2.999	2.170	0.959	0.994	1.256	0.995
Mithramycin	0.899	1.449	1.191	1.122	1.233	1.247	1.353	1.332	1.297	1.053
Monensin	0.834	1.022	1.031	0.873	2.263	1.921	0.914	1.001	0.941	0.733
Mycolactone	0.194	0.425	0.154	0.265	0.476	0.377	0.130	0.442	0.253	0.505
Myxothiazol	2.451	2.151	3.288	2.652	7.050	4.891	3.084	2.922	3.064	1.833

**Table 5.** Q values of ten non-antibiotic-producing bacteria based on blastp alignment against 15 gene clusters involved in antibiotic synthesis.

profiles, a total of 92 metabolic substances were identified by UNIFI workstation according to the mass-to-charge ratio of molecular ions and their corresponding fragment ions. Amongst them, six compounds, possessing antibacterial, antifungal or antitumor activity, were concerned (Fig. 9: compound 1–6). They affiliate to four groups: diketopiperazines (DKPs) group<sup>24</sup> (compound 1–3), phenoxazine derivatives<sup>25</sup> (compound 4), alpha-pyrone group<sup>26</sup> (compound 5), and pyranonaphthoquinone (PNQ) group<sup>27</sup> (compound 6). Concretely speaking, three diketopiperazines (DKPs) compounds (1–3) produced by a deep-sea-derived *Nocardioopsis alba* SCSIO 03039 formerly reported did not showed any antibacterial activities against four indicator strains *Escherichia coli* ATCC 25922, *Staphylococcus aureus* ATCC 29213, *Bacillus subtilis* SCSIO BS01 and *Bacillus thuringiensis* SCSIO BT01<sup>24</sup>. The antibacterial efficacy of Phenoxazine derivative (4) produced by a marine derived strain, *Nocardioopsis* sp. 236 against *M. smegmatis* was tested, yet it didn't show any antibacterial activity<sup>28</sup>. Nocardioapyrones A (5) produced by *Nocardioopsis alkaliphila*, an alkalophilic actinomycete showed weak antibacterial activities against *Pseudomonas aeruginosa*, *Enterobacter aerogenes* and *Escherichia coli* with MIC values of 20, 40 and 40 μM and showed no activity against both *Staphylococcus aureus* and *Candida albicans* with more than 100 μM of MIC values<sup>26</sup>. Griseusins G (6) produced by *Nocardioopsis* sp. YIM DT266, also an alkalophilic actinomycete, showed strong inhibitory activity against *Staphylococcus aureus* ATCC 29213, *Micrococcus luteus*, and *Bacillus subtilis*

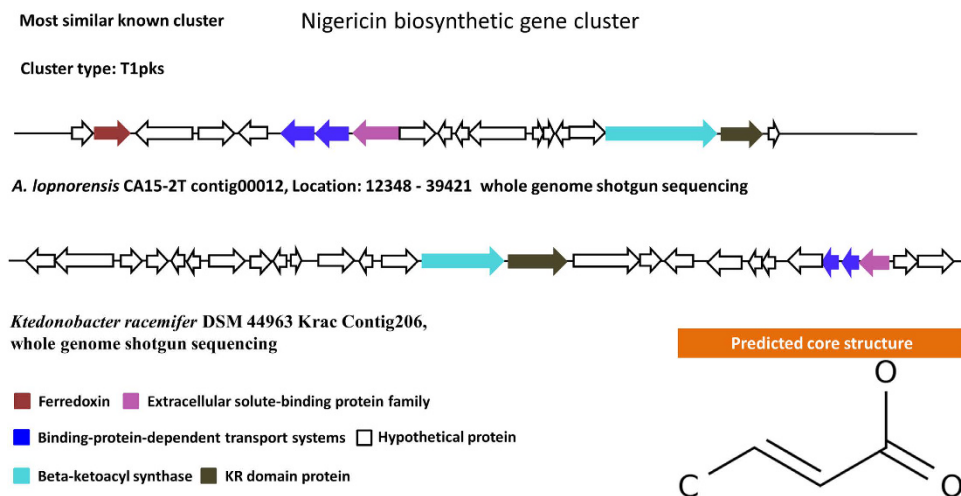
Type	Contig ID	Region	Most similar known cluster	Homology (%)	Core structure
Nrps	contig00002	146396–187281	Oxazolomycin	15	
T1pks	contig00004	27809–79280	Simocyclinone	8	
T1pks	contig00006	1–54014	Apoptolidin	23	
T1pks	contig00058	12348–39421	Nigericin	50	
T1pks	contig00127	1–7562	Aculeximycin	23	-
T1pks	contig00147	1–4638	-	-	-
T1pks	contig00161	1–2530	-	-	-
Ectoine	contig00009	1–7878	Ectoine	75	-
Siderophore	contig00003	57640–69490	Desferrioxamine B	60	-
Terpene	contig00003	127547–155465	Actinomycin	10	-
Terpene	contig00045	33564–47160	Carotenoid	27	-
Terpene	contig00148	1–3944	Carotenoid	27	-
Butyrolactone	contig00074	15307–30946	-	-	-
Other	contig00004	87089–129447	Kanamycin	3	-
Other	contig00008	50915–91946	Herboxidiene	2	-
Other	contig00023	41327–72399	Lipstatin	35	-
Other	contig00026	1–32755	-	-	-

**Table 6. Overview of 17 secondary metabolic biosynthesis gene clusters of *A. lopnorenensis* CA15-2<sup>T</sup> detected by anti SMASH2.0 server.**

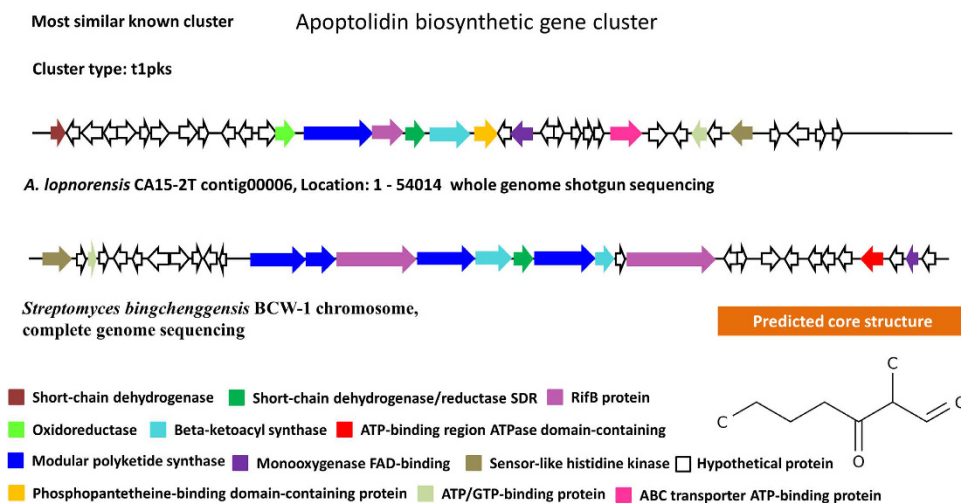


**Figure 5. Proposed biosynthetic gene cluster of simocyclinone-like compound in *A. lopnorenensis* CA15-2<sup>T</sup>.** The most similar gene cluster from *Nocardioopsis dassonvillei* is shown, with related genes depicted in the same color to highlight intracluster gene rearrangements. Predicted core structure displayed in the lower right corner.

with MIC values of 0.8, 1.47 and 1.33  $\mu\text{g/ml}$ , respectively<sup>29</sup>. Antimicrobial results of the six known compounds identified by UPLC-MS revealed that only Griseusins G contributed inhibitory activity against Gram-positive bacteria, such as *Staphylococcus aureus*, but none showed any activities against Gram-negative bacteria, such as



**Figure 6.** Proposed biosynthetic gene cluster of nigericin-like compound in *A. lopnorenensis* CA15-2<sup>T</sup>. The most similar gene cluster from *Ktedonobacter racemifer* is shown, with related genes depicted in the same color to highlight intracluster gene rearrangements. Predicted core structure displayed in the lower right corner.



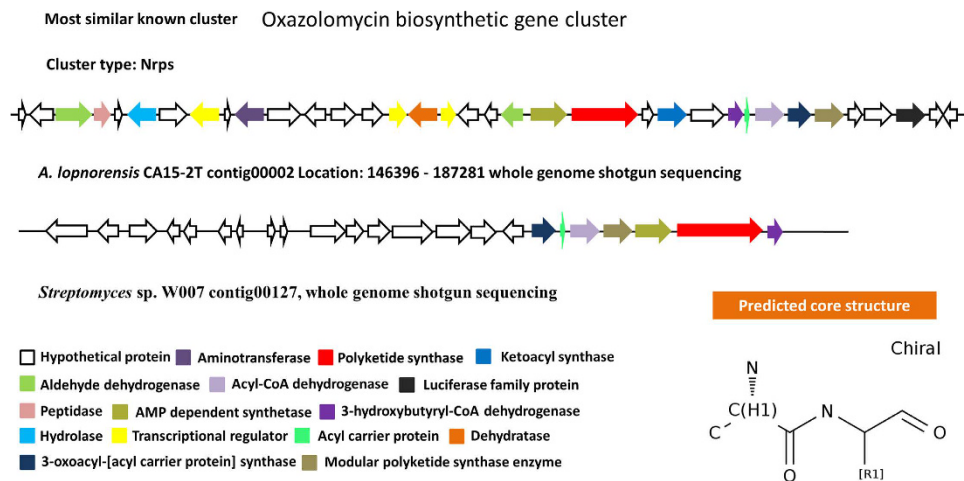
**Figure 7.** Proposed biosynthetic gene cluster of apoptolidin-like compound in *A. lopnorenensis* CA15-2<sup>T</sup>. The most similar gene cluster from *Streptomyces bingchenggensis* is shown, with related genes depicted in the same color to highlight intracluster gene rearrangements. Predicted core structure displayed in the lower right corner.

*Escherichia coli*, *Klebsiella pneumonia*, especially *Acinetobacter baumannii*, against which the biggest transparent inhibition zone was produced by the strain *A. lopnorenensis* CA15-2<sup>T</sup> (Fig. 1B,C). The marked conflict between MS elucidation and antimicrobial assay implies the existence of novel antibiotic substance.

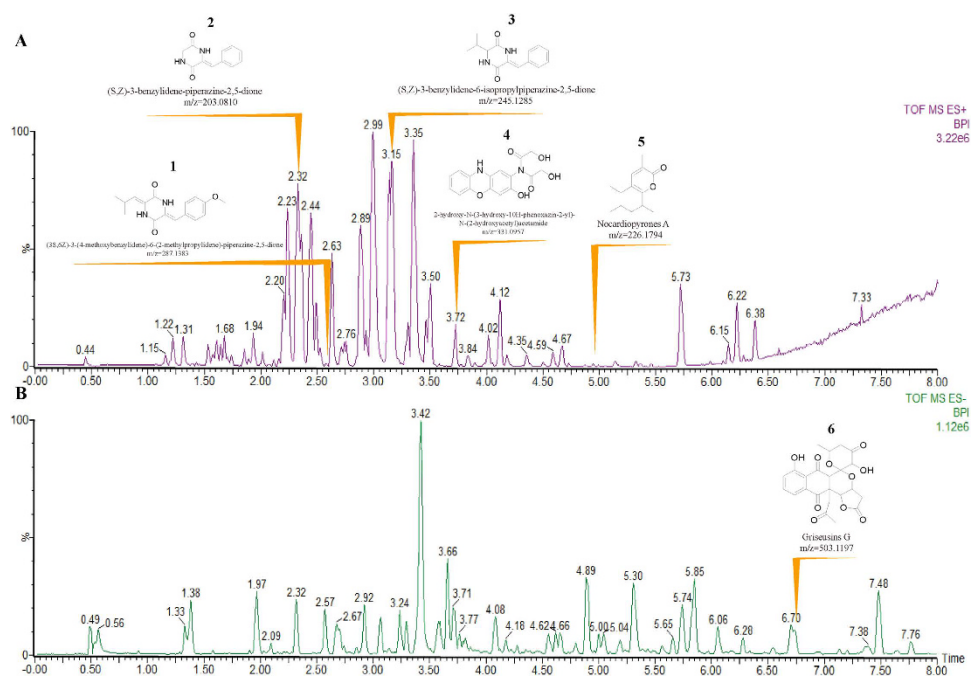
## Materials and Methods

**Strains and Growth Media.** Strain *A. lopnorenensis* CA15-2<sup>T</sup> was deposited in Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences & Peking Union Medical College and maintained on R2A (Difco, BD, USA) slants containing 5% (w/v) NaCl at 4 °C or as a 20% (v/v) glycerol suspension at –20 °C. Growth features of strain *A. lopnorenensis* CA15-2<sup>T</sup> were evaluated on R2A medium with 5% NaCl and without NaCl at 28 °C, pH 7.5 for 10 days. Both the seed medium and fermentation broth was tryptic soy broth (TSB; Bacto™, BD, USA) supplemented with 5% (w/v) NaCl. Two fungi strain, *Candida albicans* CCTCC AY93025 and *Cryptococcus laurentii* CCTCC AY91013 were obtained from China Center for Type Culture Collection in Wuhan City, Hubei Province of China and were propagated in Sabouraud Dextrose Agar (SDA, Oxoid, UK) or Sabouraud Dextrose Broth (SDB, Oxoid, UK). Eight standard strains including *Acinetobacter baumannii* ATCC 19606, *Escherichia coli* ATCC 25922, *Klebsiella pneumonia* ATCC 10031 and ATCC 700603, *Staphylococcus aureus* ATCC 25923, *Enterococcus faecalis* ATCC 33186 and ATCC 29212, *Pseudomonas aeruginosa* ATCC 27853 were obtained from American Type Culture Collection (ATCC, Rockville, MD). Four drug-resistant strains: *Acinetobacter baumannii* 2799, *Escherichia coli* 2800, *Staphylococcus aureus* 2641 and *Pseudomonas aeruginosa* 2774 were obtained from





**Figure 8.** Proposed biosynthetic gene cluster of oxazolomycin-like compound in *A. lopnorenensis* CA15-2<sup>T</sup>. The most similar gene cluster from *Streptomyces* sp. W007 is shown, with related genes depicted in the same color to highlight intracluster gene rearrangements. Predicted core structure displayed in the lower right corner.



**Figure 9.** High-resolution LC-MS analysis of ethyl acetate (EA) extract of the fermentation broth of strain *A. lopnorenensis* CA15-2<sup>T</sup>. The chromatographic separation was performed on an ACQUITY UPLC HSS T3 column (2.1 × 100 mm). Detailed characterization of separated metabolites was executed with a QToF mass spectrometer outfitted with a high transmission efficiency collision cell (Xevo G2-XS QToF, Waters, Milford, MA). Two ESI ionization modes in MS analysis, positive (A) and negative (B), were carried out respectively to give the diverse molecular properties in the EA extract. Structures and mass-to-charge ratios\* of the six antibiotics or antibiotic substances (compounds 1–6) show upon their specific retention times correspondingly. \*For the six antibiotics or antibiotic substances: compounds 1–5 in Positive ionization mode (Nocardioapyrones A: [M + NH<sub>4</sub>]<sup>+</sup>; others: [M + H]<sup>+</sup>); compound 6 in Negative ionization mode (Griseusins G: [M-HCOO]<sup>-</sup>).

Beijing Hospitals. Mueller-Hinton broth (MHB; Oxoid, UK) was used as the cultivation medium of all 12 bacteria strains.

**Antimicrobial Assay.** Antimicrobial assay were carried out by paper-disc agar diffusion method<sup>30</sup>. The stock culture of strain CA15-2<sup>T</sup> was inoculated into 250 mL Erlenmeyer flasks containing 50 mL of TSB (BD) supplemented with 5% (w/v) NaCl as seed culture medium, the seed culture was harvested after 5 days incubation at 37 °C on a rotary shaker at 180 rpm and then 50 mL of culture broth were transferred to 1 L of producing medium

with the same ingredients as seed culture medium in 5 L Erlenmeyer flasks. Four sets of 5 L Erlenmeyer flasks were incubated in the same culture condition as the seed culture. Four liters of the fermentation broth were harvested after incubation for 10 days and further centrifuged at 8000 rpm for 10 min at room temperature. Four liters of supernatant were extracted with the same volume of ethyl acetate and the organic layer was further concentrated to dryness under reduced pressure, and then finally was dissolved in 2 mL methanol as sample for test. Forty microlitres sample were loaded on sterilized filter paper with a diameter of 6 mm. After dryness, filter papers were covered on the agar plate containing different fungi or bacteria as indicator strain and antimicrobial activities were observed and recorded after 24 hours incubation at 37 °C and the zones of inhibition were measured.

**Mass Spectrometric Analysis.** Concentrated sample of EA extracts was analyzed on ACQUITY™ Ultra Performance Liquid Chromatography system (Waters, USA). The chromatographic separation was performed on a ACQUITY UPLC HSS T3 column (2.1 × 100 mm I.D., 1.8 μm). The mobile phase consisted of water containing 0.1% formic acid (A) and acetonitrile (B). The gradient condition was 1–99% B at 0–10 min, 99% B at 10–11 min, 99–1% B at 11–13 min. The flow rate was 0.45 mL/min.

Detailed characterization of each peaks separated by UPLC was performed with a Xevo™ G2-XS QToF (Waters, Manchester, UK), quadrupole and orthogonal acceleration time-of-flight tandem mass spectrometer. In negative ion mode capillary voltage was set to 2.5 kV, source temperature to 120 °C, desolvation temperature to 500 °C, and sample cone voltage to 40 V. In positive ion mode capillary voltage was set to 3.0 kV, source temperature to 120 °C, desolvation temperature to 500 °C, and sample cone voltage to 40 V. In both modes, mass spectra were acquired at a speed of 0.2 s/scan and the scanning delay of 0.01 s during analysis. In the MS<sup>E</sup> experiments, collision energy was set to scan between 6 eV and 15–45 eV. The analysis results were acquired with the lockspray to guarantee accuracy Leucine-enkephalin was used as the lockmass at a concentration of 400 ng/mL and flow rate of 10 μL/min. Data acquisition and analysis were controlled by Waters UNIFI V1.71 software. The scan range in MS and MS/MS modes were over a range of 50–1200 m/z.

**Genome Sequencing, Assembly and Annotation.** The genomic DNA library was constructed using the Ion Xpress Plus Fragment Library Kit (Life Technologies) after enzymatic digestion. The library was sequenced using the Ion Torrent platform. Since DNA fragmentation bias can lead to uneven coverage, we constructed another library with the sonication fragmentation method that was also sequenced using the Ion Torrent platform. Genome assembly of the pooled sequencing reads was performed by the Newbler program, after removing short, low-quality reads. Glimmer (ver. 3.0)<sup>31</sup> and Prodigal<sup>32</sup> were used for gene prediction. The latter was used to minimize prediction errors by Glimmer for high GC prokaryotes<sup>33</sup>. The prediction results of longer gene length were adopted if the location sites of genes overlap by the two prediction methods. Transfer RNAs (tRNAs) were predicted by tRNAscan<sup>34</sup>, and ribosomal RNAs (rRNAs) by RNAmmer<sup>35</sup>. All the putative protein-coding genes were searched against the non-redundant protein sequence (NR) database of the National Center for Biotechnology Information (NCBI) using BLASTP<sup>36</sup> for annotation. The functional category was assigned by searching against eggNOG 3.0<sup>37</sup>.

**Phylogeny and Genome Comparison.** To demonstrate a phylogenetic evolutionary relationship with other species, the complete 16S rRNA gene sequence of *A. lopnorenensis* CA15-2<sup>T</sup> was retrieved from the annotation results. Lin Guo *et al.* have demonstrated that this species is a new member of family *Nocardiopsaceae* based on phylogenetic analysis using 16S rRNA sequence<sup>22</sup>. In this study, we added nine affiliative species of *A. lopnorenensis* CA15-2<sup>T</sup> which had been deposited in the NCBI database to further verify the branches. These species are *Nocardiopsis alba* ATCC BAA 2165 (Biosample accession: SAMN02603688), *Nocardiopsis dassonvillei* DSM43111 (Biosample accession: SAMN02598425), *Thermobifida fusca* YX (Biosample accession: SAMN02598543), *Nocardioides* sp. JS614 uid58149 (Biosample accession: SAMN02598280), *Streptomyces coelicolor* A3 (2) uid57801 (Biosample accession: SAMEA1705940), *Actinobacillus succinogenes* 130Z uid58247 (Biosample accession: SAMN02598296), *Actinoplanes missouriensis* 431 uid158169, *Actinoplanes* sp. SE50 110 uid162333 (Biosample accession: SAMN02603081) and *Actinosynnemamirum* DSM 43827 uid58951 (Biosample accession: SAMN00001904). Sequence alignment was done using MUSCLE<sup>38</sup>. The aligned 16S rRNA gene sequences were used to reconstruct the phylogeny using Maximum Likelihood (ML) algorithm by MEGA<sup>39</sup>. For cases of multiple 16S rRNA gene copies in the genome, the longest sequence was chosen for phylogenetic analysis. In addition, 10 house-keeping genes (*acsA*, *aroE*, *dnaE*, *guaA*, *gyrB*, *mutL*, *ppsA*, *pyrC*, *recA* and *rpoB*)<sup>40</sup> were concatenated to reconstruct the phylogeny for comparison with the one by 16S rRNA. Three species with *A. lopnorenensis* CA15-2<sup>T</sup> in the same branch of the phylogenetic tree (Fig. 2) were used for further comparative analysis by the Ortho MCL software package<sup>41</sup>, which was used to identify ortholog, inparalog and co-ortholog groups. Finally, overlapping of the orthologous families were visualized in a Venn diagram created by the R software package.

**Identification and assessment of antibiotics-producing of gene clusters involved in secondary metabolites.** The assembled contigs of *A. lopnorenensis* CA15-2<sup>T</sup> were submitted to the anti SMASH2.0 server to search for potential secondary metabolic biosynthetic gene clusters<sup>10</sup>. Type I pks gene clusters, in which core structures were identified in anti SMASH detection were extracted for comparison with known gene clusters of other species using BLAST<sup>36</sup>. Multiple amino acid alignment of core proteins was performed by ClustalW<sup>42</sup> and depicted by boxshade ([http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)). On the other hand, in order to assess the antibiotics-producing capacity of *A. lopnorenensis* CA15-2<sup>T</sup>, 15 gene clusters involved in synthesizing the known diverse antibiotics were randomly extracted from the DOBISCUIT database<sup>43</sup> (Table S5) to be compared with the genes of *A. lopnorenensis* CA15-2<sup>T</sup> using BLAST, respectively. Furthermore, 10 gene sets from 10 known antibiotic-producing bacteria, which were picked up according to literature retrieval<sup>44–50</sup> (Table S6), were also compared with the 15 gene clusters, respectively. Ten non-antibiotics-producing bacteria were retrieved from the

NCBI database as the control group (Table S6), and the corresponding gene sets of these bacteria were also compared with the 15 gene clusters, respectively. It should be noted that none of research investigated which bacteria could not produce antibiotics, in such case; we select 'specific' bacteria as non-antibiotics-producing bacteria which might not be expected to produce antibiotics, e.g. pathogenic bacteria, acid-tolerant, halo tolerant or heat-resistant bacteria. In order to evaluate the antibiotics-producing capacity of these ten antibiotic-producing bacteria and ten non-antibiotics-producing bacteria, as well as *A. lopnorenensis* CA15-2<sup>T</sup>, index Q value was developed, in which the identity and alignment length were taken into account, and gene amount was used for normalization, defined as:

$$Q_{ij} = \frac{\sum_{j=1}^{N_i} I_j \times L_j}{N_i} / P_j \quad (1)$$

where  $I_j$  is the best-matched identity value to the blast result (jth species compared with the ith gene cluster),  $L$  is the best-matched alignment length of the blast result, and  $N_i$  is the gene number of the ith gene cluster and  $P_j$  is the gene amount of the jth species. The larger  $Q$ , the greater probability of the jth species contains the ith gene cluster. Moreover, Repeated Measures ANOVA test and Principal Component Analysis was used to test the significance of difference between antibiotics-producing group and non-antibiotics-producing group.

## Conclusion

Combining genome-scale physiologic, genetic, and metabolic investigation approaches will expand our knowledge pertaining to the exploitation of novel natural active compounds, and may help us to establish strategies to develop these natural products for human use. In this study, we sequenced the genome of a novel actinobacteria strain, *A. lopnorenensis* isolated from the rhizosphere of Tamarisk in the Lop Nor region, an extreme environment with hot, drought, and high salinity conditions. From the secondary metabolites predicted with the assembled genome, we identified 17 candidate gene clusters, of which 4 are novel; the remaining 13 clusters refer to type I polyketide synthase (PKS I) systems, nonribosomal peptide synthetases (NRPS) and ectoine. Regarding the known gene clusters, the majority of the genes, and several new members, were found in clusters of the 13 detectable secondary metabolites. Some genes' loci and directions changed significantly compared to those in reported clusters, and some new members joined these clusters. The findings reflect the flexibility and evolutionary diversity of clusters of secondary metabolites, and also bring new insights for future research on the secondary metabolites of actinomycetes.

## Discussion

Bacteria are rich resources for the production of secondary metabolites, which represent a great diversity of economically important chemicals. Actinomycetes, one of the most-important bacterial species, have been, and continue to be, a leading source for new drug discovery<sup>1</sup>. In recent years, the majority of actinomycetes-based screening programs at large pharmaceutical companies encountered hurdles due to high costs and frequent re-discovery of known compounds, leading to the redundancy<sup>6</sup>. The discovery of new strains, from extreme environments, has become a hot topic because it provides opportunities for the discovery of novel chemical structures and scaffolds with new biological functions. In the present study, *A. lopnorenensis* CA15-2<sup>T</sup> was isolated from the Lop Nor region. Lop Nor, which translates to "Lop Lake" in Mongolian, was a former salt lake, and is now completely dried-up. Few forms of life, except some special microbial populations, can exist in such extreme environment<sup>51,52</sup>. Culture experiment shows that *A. lopnorenensis* CA15-2<sup>T</sup> grew well in the medium with 5% NaCl. These microorganisms have evolved into halophilic bacteria and archaea with the development of hypersaline habitats, e.g. evaporation of marine saltwater. Generally speaking, microorganisms of this group may accumulate highly water-soluble organic compounds, which are often named 'compatible solutes'<sup>53</sup>, to maintain osmotic equilibrium with surrounding medium; thus, they can keep their cytoplasm and the cell's interior remains basically unchanged. The majority of compatible solutes are the amino acid derivatives glycine-betaine and ectoine in halophilic bacteria<sup>52</sup>. In this study, ectoine biosynthetic gene cluster was detected using whole genome sequencing. The metabolic pathway of Ectoine had been worked out by previous studies<sup>54,55</sup>. Briefly, aspartate-semialdehyde is transaminated to 2,4-diaminobutyric acid (DABA) by catalyzing with DABA transaminase (EctB), and DABA is then transferred to N $\gamma$ -acetyl-L-2,4-diaminobutyric by DABA-N $\gamma$ -acetyltransferase (EctA). Finally, ectoine synthase (EctC) catalyzes the cyclic condensation of N $\gamma$ -acetyl-L-2, 4-diaminobutyric acid, which leads to the formation of ectoine. Ectoine hydroxylase (EctD) can convert some of the ectoine to 5-hydroxyectoine under certain stress conditions (Figure S11). The identification of Ect ABCD genes based on deep sequencing facilitate us to understand the adaptation of *A. lopnorenensis* CA15-2<sup>T</sup> to a salty environment.

Concerning antimicrobial experiments, *A. lopnorenensis* CA15-2<sup>T</sup> showed a wide-ranging antimicrobial activity, in particular effectively inhibiting the growth of *Acinetobacter baumannii*, which provides clues to the possible presence of secondary metabolites as antibacterial substances. *Acinetobacter baumannii* belonging to Nonfermentative Gram-negative bacilli, has aroused wide concern in clinical and microbiology research, due to the spread of multidrug-resistant (MDR) bacillary strains among critically ill, hospitalized patients, sometimes in form of epidemics<sup>56,57</sup>. Multidrug-resistant *Acinetobacter baumannii* has been recognized to be among the most difficult to treat gram-negative bacilli<sup>58</sup>. Hence, discovery of novel antibiotics in *A. lopnorenensis* CA15-2<sup>T</sup> undoubtedly brings new hope for fighting against *Acinetobacter baumannii*. Mass spectrometric analysis reveals 6 compounds with antifungal or antitumor activity. However, only Griseusins G among these compounds demonstrated inhibitory activity against Gram-positive bacteria (i.e., *Staphylococcus aureus*). None of them showed any activities against Gram-negative bacteria (i.e., *Acinetobacter baumannii*). In contrast to the antibiotic action merely against Gram-positive bacteria by known compound, findings from antimicrobial assay may infer a novel, yet antibiotic-related gene pool, and compounds encoded by them, in *Allosalimacinospora lopnorenensis* CA15-2<sup>T</sup>.

Normally, secondary metabolites of a microorganism are catalyzed by a series of enzyme-encoding genes that are usually organized and located in a specific section of the genome, also known as gene clusters. This feature allows us to identify potential biosynthetic clusters based on genome mining approach, e.g. basic homology sequence search BALST against existing gene clusters. To date, the genome-guided approach has been used in studies of microbial natural products<sup>4</sup>. Complete genome sequencing has provided unparalleled access to secondary metabolite producing genes, especially using high-throughput sequencing techniques, which have been widely used to successfully discover natural products in microorganisms for several years<sup>1,4</sup>. Furthermore, whole genome analysis in microorganisms is increasingly being performed by diverse research, clinical and public health laboratory groups<sup>59</sup>. In this study, the genome of *A. lopnorenensis* CA15-2<sup>T</sup> was sequenced using the Ion Torrent PGM platform. Two different DNA fragmentation methods were used to construct a DNA library. Although high G + C content, multicopy segments in the genome and sequencing errors might cause frequent termination of contig assembly, which is not easily filled by conventional sequencing methods, two sequences could provide sufficient sequence and complementary information for further analysis. Through biosynthetic cluster detection, 17 known or predicted gene clusters were figured out, majority of clusters were involved in production of antibiotic substances, including several well-known compounds synthetic pathways like simocyclinone, apoptolidin and oxazolomycin. Concretely speaking, simocyclinone is active against gram-positive bacteria and some studies have showed that it may process distinct cytostatic activities against human tumor cell lines<sup>60,61</sup>. In addition, Ruth H. Flatman *et al.* demonstrated that it may inhibit an early step of the gyrase catalytic cycle by preventing binding of the enzyme to DNA<sup>62</sup>. Apoptolidin has been proved to be a selective cytotoxic macrolide compound<sup>63</sup>, which can be produced by the polyketide synthesis gene cluster of *Streptomyces bingchenggensis* BCW-1. Some studies showed that apoptolidin exhibit selective activity against cells, and it could induce cell death in malignant cell lines when applied together with LDH inhibitor oxamate<sup>64,65</sup>. Furthermore, oxazolomycin that might be synthesized by NRPS cluster in *A. lopnorenensis* CA15-2<sup>T</sup>, has been proven to show diverse and important antibacterial, antitumor and anti-human immunodeficiency virus activity<sup>66</sup>.

Remarkably, the genes' loci and directions changed, and some orphan genes were found in some of these clusters, which implies *A. lopnorenensis* CA15-2<sup>T</sup> probably produces novel antibiotics. Given the transcriptional status and common co-existence of transcriptional regulators in gene clusters of secondary metabolites, the loci of genes might influence the final assembly of the secondary metabolic complex, because different transcription factors may regulate the expression level of genes. Likewise, the direction of genes may regulate expression in transcription levels. Moreover, metabolic substance's gene cluster might be silenced by a specific regulation mechanism in genome<sup>67</sup>, and just expressed under particular conditions, which might be one of reasons that we could not identify the specific products from MS analysis. Another probably reason is that *A. lopnorenensis* CA15-2<sup>T</sup> has the capacity of producing other classes of antimicrobial substances, such as antimicrobial peptides, which could not be detected by MS analysis. Actually, another major class of antimicrobial agents is composed of ribosomally synthesized and post-translationally modified peptides (RiPPs), which serve as a promising alternative to conventional antibiotics biosynthesized by polyketides or non-ribosomal pathways<sup>68,69</sup>. In order to test this hypothesis, genome scanning was carried out using BAGEL3 server<sup>70</sup>. The results revealed that *A. lopnorenensis* CA15-2<sup>T</sup> genome might encode two RiPPs: Linear azol(in)e-containing peptide (LAP) (Figure S12) and Lasso peptide (Figure S13). LAPs have been widely identified by mining publicly available genome sequences. According to previous studies<sup>68,69</sup>, the critical components in LAP biosynthesis are the inactive precursor peptide and the heterotrimeric synthetase complex comprised of a dehydrogenase and cyclodehydratase, all of which were identified in *A. lopnorenensis* CA15-2<sup>T</sup> (Figure S12). In addition, some LAPs have been proven to exhibit antibiotic activity, e.g. Goadsproin produced by *Streptomyces* sp. TPA0584<sup>69</sup>. Lasso peptides, the most extraordinary RiPPs, are regarded as promising antibacterial peptides due to their specific knotted structure<sup>71</sup>, the lasso fold that gives them stability against heat, chemical attack and proteases<sup>72</sup>. Previous studies<sup>69</sup> on the biosynthesis of Lasso peptide have revealed that at least four encoding genes are essential for its formation. Three of them were detected in the genome of *A. lopnorenensis* CA15-2<sup>T</sup> except for the gene encoding transporter (Figure S13), which may transport the lasso peptide out of cell. One possible reason is that incomplete genome assembly of *A. lopnorenensis* CA15-2<sup>T</sup> leading to the interruption of gene cluster, or the gene of *A. lopnorenensis* CA15-2<sup>T</sup> is so distinct with the existing ones that we could not identify it based on sequence similarity search. Majority of Lasso peptides are enzyme inhibitors or receptor antagonists, more importantly, some of them have been proven to possess antibacterial activity against either Gram-negative bacteria or Gram-positive bacteria<sup>69</sup>. This finding may provide the clue to interpret that *A. lopnorenensis* CA15-2<sup>T</sup> has the capability of inhibiting against *Acinetobacter baumannii*. Certainly, further experimental investigation is requisite to validate the truth of these RiPPs in *A. lopnorenensis* CA15-2<sup>T</sup>, e.g. proteomics approach. In conclusion, this study demonstrated that *A. lopnorenensis* CA15-2<sup>T</sup> has enormous potential to antibiotic development. Despite efforts and progress made towards gene function identification in the field of analytical chemistry, it's still not easy to connect the detected genotypes with chemotypes, indicating the identification of these metabolites remains a big challenge.

## References

- McAlpine, J. B. *et al.* Microbial genomics as a guide to drug discovery and structural elucidation: ECO-02301, a novel antifungal agent, as an example. *J Nat Prod* **68**, 493–496 (2005).
- Berdy, J. Bioactive microbial metabolites. *J Antibiot (Tokyo)* **58**, 1–26 (2005).
- Gontang, E. A., Gaudencio, S. P., Fenical, W. & Jensen, P. R. Sequence-based analysis of secondary-metabolite biosynthesis in marine actinobacteria. *Appl Environ Microbiol* **76**, 2487–2499 (2010).
- Udwary, D. W. *et al.* Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc Natl Acad Sci USA* **104**, 10376–10381 (2007).
- Hong, K. *et al.* Actinomycetes for marine drug discovery isolated from mangrove soils and plants in China. *Mar Drugs* **7**, 24–44 (2009).
- Zotchev, S. B. Marine actinomycetes as an emerging resource for the drug development pipelines. *J Biotechnol* **158**, 168–175 (2012).

7. Subramani, R. & Aalbersberg, W. Marine actinomycetes: an ongoing source of novel bioactive metabolites. *Microbiol Res* **167**, 571–580 (2012).
8. Ito, T. [Biosynthetic study of actinomycetes-metabolites for creating novel analogs]. *Yakugaku Zasshi* **133**, 1007–1015 (2013).
9. Yassin, A. & Mankin, A. S. Potential new antibiotic sites in the ribosome revealed by deleterious mutations in RNA of the large ribosomal subunit. *J Biol Chem* **282**, 24329–24342 (2007).
10. Blin, K. *et al.* antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res* **41**, W204–212 (2013).
11. Gagne, J. J. & Choudhry, N. K. How many “me-too” drugs is too many? *Jama* **305**, 711–712 (2011).
12. Hong, K. [Actinomycetes from mangrove and their secondary metabolites]. *Wei Sheng Wu Xue Bao* **53**, 1131–1141 (2013).
13. Shi, Y., Zhang, X. & Lou, K. [Endophyte microbial community in *Achnatherum inebrians*]. *Wei Sheng Wu Xue Bao* **52**, 1297–1308 (2012).
14. Subramani, R., Kumar, R., Prasad, P., Aalbersberg, W. & Rethesh, S. T. Cytotoxic and antibacterial substances against multi-drug resistant pathogens from marine sponge symbiont: Citrinin, a secondary metabolite of *Penicillium* sp. *Asian Pac J Trop Biomed* **3**, 291–296 (2013).
15. Ballav, S., Kerkar, S., Thomas, S. & Augustine, N. Halophilic and halotolerant actinomycetes from a marine saltern of Goa, India producing anti-bacterial metabolites. *Journal of bioscience and bioengineering* **119**, 323–330 (2015).
16. Schroeckh, V. *et al.* Intimate bacterial-fungal interaction triggers biosynthesis of archetypal polyketides in *Aspergillus nidulans*. *Proc Natl Acad Sci USA* **106**, 14558–14563 (2009).
17. Gross, H. Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. *Appl Microbiol Biotechnol* **75**, 267–277 (2007).
18. Hu, Y., Chou, W. K., Hopson, R. & Cane, D. E. Genome mining in *Streptomyces clavuligerus*: expression and biochemical characterization of two new cryptic sesquiterpene synthases. *Chem Biol* **18**, 32–37 (2011).
19. Soror, S. H., Rao, R. & Cullum, J. Mining the genome sequence for novel enzyme activity: characterisation of an unusual member of the hormone-sensitive lipase family of esterases from the genome of *Streptomyces coelicolor* A3 (2). *Protein Eng Des Sel* **22**, 333–339 (2009).
20. Lautru, S., Deeth, R. J., Bailey, L. M. & Challis, G. L. Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. *Nat Chem Biol* **1**, 265–269 (2005).
21. Liu, B. *et al.* [Biodiversity and functional enzymes of cultured halophilic archaeon in Lop Nur region]. *Wei Sheng Wu Xue Bao* **51**, 1222–1231 (2011).
22. Guo, L. *et al.* *Allosalinactinospira lopnorenensis* gen. nov., sp. nov., a new member of the family Nocardiospiceae isolated from soil. *Int J Syst Evol Microbiol* **65**, 206–213 (2015).
23. Schwibbert, K. *et al.* A blueprint of ectoine metabolism from the genome of the industrial producer *Halomonas elongata* DSM 2581 T. *Environ Microbiol* **13**, 1973–1994 (2011).
24. Zhang, Q. *et al.* New diketopiperazine derivatives from a deep-sea-derived *Nocardiopsis alba* SCSIO 03039. *J Antibiot (Tokyo)* **66**, 31–36 (2013).
25. Imai, S., Shimazu, A., Furihata, K., Hayakawa, Y. & Seto, H. Isolation and structure of a new phenoxazine antibiotic, exfoliazone, produced by *Streptomyces exfoliatus*. *J Antibiot (Tokyo)* **43**, 1606–1607 (1990).
26. Wang, Z. *et al.* New pyran-2-ones from alkaliphilic actinomycete, *Nocardiopsis alkaliphila* sp. Nov. YIM-80379. *Chem Biodivers* **10**, 281–287 (2013).
27. Brimble, M. A., Duncalf, L. J. & Nairn, M. R. Pyranonaphthoquinone antibiotics-isolation, structure and biological activity. *Nat Prod Rep* **16**, 267–281 (1999).
28. Lu, C. H., Li, Y. Y., Wang, H. X., Wang, B. M. & Shen, Y. M. A new phenoxazine derivative isolated from marine sediment actinomycetes, *Nocardiopsis* sp. 236. *Drug Discov Ther* **7**, 101–104 (2013).
29. Ding, Z. G. *et al.* Griseusins F and G, spiro-naphthoquinones from a tin mine tailings-derived alkaliphilic *Nocardiopsis* species. *J Nat Prod* **75**, 1994–1998 (2012).
30. Murray, P. R. *et al.* Manual of Clinical Microbiology (6th edn). *Trends in Microbiology* **3**, 449–449 (1995).
31. Delcher, A. L., Harmon, D., Kasif, S., White, O. & Salzberg, S. L. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* **27**, 4636–4641 (1999).
32. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics* **11**, 119 (2010).
33. Angelova, M. Computational Methods for Gene Finding in Prokaryotes. *ICT Innovations* 1857–7288 (2010).
34. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955–964 (1997).
35. Lagesen, K. *et al.* RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**, 3100–3108 (2007).
36. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of molecular biology* **215**, 403–410 (1990).
37. Powell, S. *et al.* eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res* **40**, D284–289 (2012).
38. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792–1797 (2004).
39. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* **28**, 2731–2739 (2011).
40. Loper, J. E. *et al.* Comparative genomics of plant-associated *Pseudomonas* spp.: insights into diversity and inheritance of traits involved in multitrophic interactions. *PLoS genetics* **8**, e1002784 (2012).
41. Li, L., Stoeckert, C. J., Jr. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* **13**, 2178–2189 (2003).
42. Thompson, J. D., Gibson, T. & Higgins, D. G. Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics*, 2.3. 1–2.3. 22 (2002).
43. Ichikawa, N. *et al.* DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. *Nucleic acids research* **41**, D408–414 (2013).
44. Watanabe, K., Okuda, T., Yokose, K., Furumai, T. & Maruyama, H. B. *Actinosynnema mirum*, a new producer of nocardicin antibiotics. *J Antibiot (Tokyo)* **36**, 321–324 (1983).
45. Beck, Z. Q., Burr, D. A. & Sherman, D. H. Characterization of the beta-methylaspartate-alpha-decarboxylase (CrpG) from the cryptophycin biosynthetic pathway. *Chembiochem: a European journal of chemical biology* **8**, 1373–1375 (2007).
46. Carter, G. T., Nietsche, J. A., Williams, D. R. & Borders, D. B. Citreamicins, novel antibiotics from *Micromonospora citrea*: isolation, characterization, and structure determination. *J Antibiot (Tokyo)* **43**, 504–512 (1990).
47. Coutanceau, E. *et al.* Modulation of the host immune response by a transient intracellular stage of *Mycobacterium ulcerans*: the contribution of endogenous mycolactone toxin. *Cellular microbiology* **7**, 1187–1196 (2005).
48. Magarvey, N. A. *et al.* Biosynthetic characterization and chemoenzymatic assembly of the cryptophycins. Potent anticancer agents from cyanobionts. *ACS chemical biology* **1**, 766–779 (2006).

49. Meier, J. L., Barrows-Yano, T., Foley, T. L., Wike, C. L. & Burkart, M. D. The unusual macrocycle forming thioesterase of mycolactone. *Mol Biosyst* **4**, 663–671 (2008).
50. Stinear, T. P. *et al.* Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *Proc Natl Acad Sci USA* **101**, 1345–1349 (2004).
51. Oren, A. Diversity of halophilic microorganisms: environments, phylogeny, physiology, and applications. *J Ind Microbiol Biotechnol* **28**, 56–63 (2002).
52. Oren, A. Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Systems* **4**, 2 (2008).
53. Brown, A. D. Microbial water stress. *Bacteriol Rev* **40**, 803–846 (1976).
54. Canovas, D., Vargas, C., Calderon, M. L., Ventosa, A. & Nieto, J. J. Characterization of the genes for the biosynthesis of the compatible solute ectoine in the moderately halophilic bacterium *Halomonas elongata* DSM 3043. *Syst Appl Microbiol* **21**, 487–497 (1998).
55. Goller, K., Ofer, A. & Galinski, E. A. Construction and characterization of a NaCl-sensitive mutant of *Halomonas elongata* impaired in ectoine biosynthesis. *FEMS Microbiol Lett* **161**, 293–300 (1998).
56. Dijkshoorn, L., Nemeč, A. & Seifert, H. An increasing threat in hospitals: multidrug-resistant *Acinetobacter baumannii*. *Nat Rev Microbiol* **5**, 939–951 (2007).
57. Maragakis, L. L. & Perl, T. M. *Acinetobacter baumannii*: epidemiology, antimicrobial resistance, and treatment options. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* **46**, 1254–1263 (2008).
58. Bergogne-Berezin, E. Treatment of *Acinetobacter* infections. *Expert opinion on investigational drugs* **6**, 119–127 (1997).
59. Edwards, D. J. & Holt, K. E. Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. *Microb Inform Exp* **3**, 2 (2013).
60. Trefzer, A. *et al.* Biosynthetic gene cluster of simocyclinone, a natural multihybrid antibiotic. *Antimicrob Agents Chemother* **46**, 1174–1182 (2002).
61. Schimana, J., Walker, M., Zeeck, A. & Fiedler, P. Simocyclinones: diversity of metabolites is dependent on fermentation conditions. *J Ind Microbiol Biotechnol* **27**, 144–148 (2001).
62. Flatman, R. H., Howells, A. J., Heide, L., Fiedler, H. P. & Maxwell, A. Simocyclinone D8, an inhibitor of DNA gyrase with a novel mode of action. *Antimicrob Agents Chemother* **49**, 1093–1100 (2005).
63. Salomon, A. R., Voehringer, D. W., Herzenberg, L. A. & Khosla, C. Apoptolidin, a selective cytotoxic agent, is an inhibitor of F0F1-ATPase. *Chem Biol* **8**, 71–80 (2001).
64. Diogo, C. V. *et al.* Berberine as a promising safe anti-cancer agent - is there a role for mitochondria? *Current drug targets* **12**, 850–859 (2011).
65. Wender, P. A. & Longcore, K. E. Isolation, structure determination, and anti-cancer activity of apoptolidin D. *Organic letters* **9**, 691–694 (2007).
66. Zhao, C. *et al.* Oxazolomycin biosynthesis in *Streptomyces albus* JA3453 featuring an “acyltransferase-less” type I polyketide synthase that incorporates two distinct extender units. *J Biol Chem* **285**, 20097–20108 (2010).
67. Yamanaka, K. *et al.* Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc Natl Acad Sci USA* **111**, 1957–1962 (2014).
68. Letzel, A. C., Pidot, S. J. & Hertweck, C. Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) in anaerobic bacteria. *BMC genomics* **15**, 983 (2014).
69. Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Natural product reports* **30**, 108–160 (2013).
70. van Heel, A. J., de Jong, A., Montalban-Lopez, M., Kok, J. & Kuipers, O. P. BAGEL3: Automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res* **41**, W448–453 (2013).
71. Maksimov, M. O., Pelczar, I. & Link, A. J. Precursor-centric genome-mining approach for lasso peptide discovery. *Proc Natl Acad Sci USA* **109**, 15223–15228 (2012).
72. Maksimov, M. O. & Link, A. J. Prospecting genomes for lasso peptides. *J Ind Microbiol Biotechnol* **41**, 333–344 (2014).

## Acknowledgements

We greatly thank Xu Bai from Waters Corporation for his advices in MS analysis. This study was partly supported by grants from the Science and Technology Development Fund (FDCT) of Macao SAR (Ref. No. 134/2014/A3), Research Committee of University of Macau (MYRG139(Y1-L4)-ICMS12-LMY and MYRG2015-00214-ICMS-QRCM), and the Overseas and Hong Kong, Macau Young Scholars Collaborative Research Fund by the Natural National Science Foundation of China (Grant no. 81328025). Meanwhile, this study was partly supported by grants from National Natural Sciences Foundation of China (NSFC, Grant no. 81373308 and Grant no. 81321004).

## Author Contributions

S.M.Y.L. and C.H.S. conceived and designed the study. L.T. and L.G. performed the experiments. C.H. and R.K.K.L. analyzed the data. C.H. and R.K.K.L. wrote the paper. S.M.Y.L., C.H.S., M.G., Y.W.W. and I.L. reviewed and edited the manuscript. All authors read and approved the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Huang, C. *et al.* Genome-guided Investigation of Antibiotic Substances produced by *Allosalinactinospora lopnorenensis* CA15-2<sup>T</sup> from Lop Nor region, China. *Sci. Rep.* **6**, 20667; doi: 10.1038/srep20667 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>