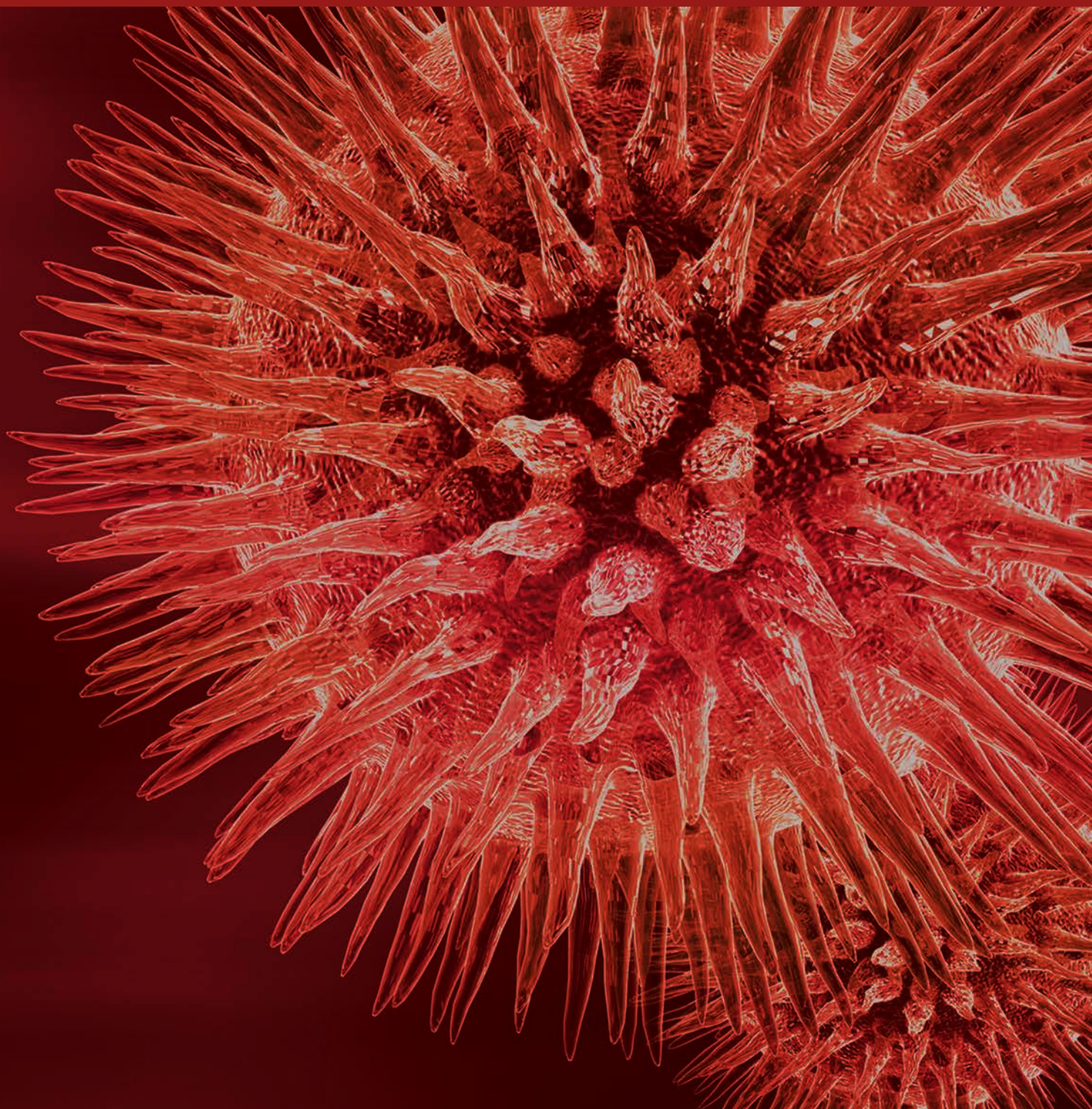


BioMed Research International

Molecular Phylogenetics 2016

Guest Editors: Vassily Lyubetsky, William H. Piel, and Peter F. Stadler





Molecular Phylogenetics 2016

BioMed Research International

Molecular Phylogenetics 2016

Guest Editors: Vassily Lyubetsky, William H. Piel,
and Peter F. Stadler



Copyright © 2016 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in “BioMed Research International.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Contents

Molecular Phylogenetics 2016

Vassily Lyubetsky, William H. Piel, and Peter F. Stadler
Volume 2016, Article ID 9029306, 2 pages

***Burkholderia contaminans* Biofilm Regulating Operon and Its Distribution in Bacterial Genomes**

Olga L. Voronina, Marina S. Kunda, Natalia N. Ryzhova, Ekaterina I. Aksenova, Andrey N. Semenov, Yulia M. Romanova, and Alexandr L. Gintsburg
Volume 2016, Article ID 6560534, 13 pages

Resistance of Permafrost and Modern *Acinetobacter lwoffii* Strains to Heavy Metals and Arsenic Revealed by Genome Analysis

Sofia Mindlin, Anatolii Petrenko, Anton Kurakov, Alexey Beletsky, Andrey Mardanov, and Mayya Petrova
Volume 2016, Article ID 3970831, 9 pages

Molecular Cloning and Characterization of Full-Length cDNA of Calmodulin Gene from Pacific Oyster *Crassostrea gigas*

Xing-Xia Li, Wen-Chao Yu, Zhong-Qiang Cai, Cheng He, Na Wei, Xiao-Tong Wang, and Xi-Qing Yue
Volume 2016, Article ID 5986519, 7 pages

Phylogeographic Structure of a Tethyan Relict *Capparis spinosa* (Capparaceae) Traces Pleistocene Geologic and Climatic Changes in the Western Himalayas, Tianshan Mountains, and Adjacent Desert Regions

Qian Wang, Ming-Li Zhang, and Lin-Ke Yin
Volume 2016, Article ID 5792708, 13 pages

Reconstructing the Phylogeny of *Capsosiphon fulvescens* (Ulotrichales, Chlorophyta) from Korea Based on *rbcL* and 18S rDNA Sequences

Sang-Mi Sun, Seung Hwan Yang, Kirill S. Golokhvast, Bao Le, and Gyuhwa Chung
Volume 2016, Article ID 1462916, 6 pages

Molecular Typing Characteristic and Drug Susceptibility Analysis of *Mycobacterium tuberculosis* Isolates from Zigong, China

Hai-Can Liu, Jian-Ping Deng, Hai-Yan Dong, Ti-Quan Xiao, Xiu-Qin Zhao, Zheng-Dong Zhang, Yi Jiang, Zhi-Guang Liu, Qun Li, and Kang-Lin Wan
Volume 2016, Article ID 6790985, 7 pages

Editorial

Molecular Phylogenetics 2016

Vassily Lyubetsky,¹ William H. Piel,² and Peter F. Stadler³

¹*Russian Academy of Sciences, Moscow, Russia*

²*Yale-NUS College and National University of Singapore, Singapore*

³*Bioinformatics, Institute for Informatics, Leipzig University, Leipzig, Germany*

Correspondence should be addressed to Vassily Lyubetsky; lyubetsk@iitp.ru

Received 5 December 2016; Accepted 12 December 2016

Copyright © 2016 Vassily Lyubetsky et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Knowledge of phylogeny is of fundamental importance for understanding evolution. It has become an indispensable tool in modern genomics as a framework for interpreting genomes and metagenomes, for understanding the evolution of genes, proteins, and noncoding RNAs, as well as gene regulation by secondary RNA and protein structures, or for reconstructing ancestral genomes [1]. The era of next-generation sequencing (NGS) brought about an influx of data but also posed new theoretical challenges, for example, in reducing systematic error, insuring gene orthology, and working with incomplete datasets [2]. The contents of the special issue exemplify the wide range of uses for phylogenetics: traditional medicines, climate change, functional genomics, and microbial resistance to heavy metals and drugs.

Some topics of modern phylogenetics are to be mentioned. Traditionally, studies of species evolution to a large extent relied on the comparative analysis of genomic regions coding for rRNAs and proteins apart from the analysis of morphological characters. Later, analyses made use of regulatory elements and the structure of the genome as a whole. More recently, phylogenetic analyses are incorporating ultraconserved elements (UCEs) and highly conserved elements (HCEs). Models of evolution of the genome structure and HCE initially faced considerable algorithmic challenges, which gave rise to (often unnatural) constraints in these models even for conceptually simple tasks such as the calculation of distance between two structures or the identification of UCEs. These constraints are now being addressed with fast and efficient solutions with no constraints on the underlying models [3, 4]. These approaches have led to

an unexpected result: at least for some organelles and taxa, the genome and HCE structures, despite themselves containing relatively little information, still adequately resolve the evolution of species. The HCEs identification is also important in searching for promoters and regulatory elements that characterize the functional evolution of the genome.

Another fundamental question is the resolution of ancient taxa with obscure and recalcitrant relationships. A classic example is the question of monophyly of the Mesozoa, specifically with respect to the parasitic phyla Orthonectida and Dicyemida. This question is aggravated by a well-known and yet still unsolved problem of long branch attraction. Of particular interest is the statistical view on such questions that leads to the problem of a formal description of the classes of trees for a given supermatrix which are generated by popular programs such as PhyloBayes and RAxML. Also of interest is the development of statistical tests for monophyly within this framework which retain accuracy despite increasingly large, genome-scale, datasets.

Using regulatory elements for phylogeny is a complex problem. A key question is how to estimate statistical support for phylogenetic signal derived from regulatory elements that are highly dynamic and not easily aligned. Even simple computation of distances between genes, leader genes, or hairpins in RNA can be nontrivial [5]. An alternative approach is to extract phylogenetic signal from syntenic patterns of regulatory elements [6], but this comes with its own computational challenges.

Apart from classic molecular systematic applications to infer taxon phylogenies, the trend is to approach molecular

and biodiversity assessment at different levels in various communities, for example, at the intraspecific level and with environmental samples, including systematic studies of bacterial and viral pathogenic agents. Molecular markers such as mobile elements are being developed and exploited in studies of population polymorphisms, and RNA secondary structures are used to detect signatures of selection.

A historical profile of molecular phylogenetics with some extrapolations into the future, as well as a brief outline of hot spots in this field, can be found in this special issue [7]. We truly hope that these contributions will be of use to scientists in various areas in possibly helping them to find answers and pose new questions in their own research.

Acknowledgments

We gratefully thank all scientific reviewers who dedicated their time to evaluate submissions in this special issue.

Vassily Lyubetsky
William H. Piel
Peter F. Stadler

References

- [1] E. V. Koonin, *The Logic of Chance: The Nature and Origin of Biological Evolution*, FT Press, 2011.
- [2] J. W. Wägele and T. Bartolomaeus, Eds., *Deep Metazoan Phylogeny: The Backbone of the Tree of Life: New Insights from Analyses of Molecules, Morphology, and Theory of Data Analysis*, Walter de Gruyter, 2014.
- [3] L. I. Rubanov, A. V. Seliverstov, O. A. Zverkov, and V. A. Lyubetsky, "A method for identification of highly conserved elements and evolutionary analysis of superphylum Alveolata," *BMC Bioinformatics*, vol. 17, article 385, 2016.
- [4] V. Lyubetsky, R. Gershgorin, A. Seliverstov, and K. Gorbunov, "Algorithms for reconstruction of chromosomal structures," *BMC Bioinformatics*, vol. 17, article 40, 2016.
- [5] S. A. Korolev, O. A. Zverkov, A. V. Seliverstov, and V. A. Lyubetsky, "Ribosome reinitiation at leader peptides increases translation of bacterial proteins," *Biology Direct*, vol. 11, article no. 20, 2016.
- [6] O. A. Zverkov, A. V. Seliverstov, and V. A. Lyubetsky, "A database of plastid protein families from red algae and apicomplexa and expression regulation of the moeB Gene," *BioMed Research International*, vol. 2015, Article ID 510598, 5 pages, 2015.
- [7] V. Lyubetsky, W. H. Piel, and D. Quandt, "Current advances in molecular phylogenetics," *BioMed Research International*, vol. 2014, Article ID 596746, 2 pages, 2014.

Research Article

***Burkholderia contaminans* Biofilm Regulating Operon and Its Distribution in Bacterial Genomes**

Olga L. Voronina,¹ Marina S. Kunda,¹ Natalia N. Ryzhova,¹ Ekaterina I. Aksenova,¹ Andrey N. Semenov,¹ Yulia M. Romanova,^{1,2} and Alexandr L. Gintsburg^{1,2}

¹N.F. Gamaleya Federal Research Center for Epidemiology and Microbiology, Ministry of Health of Russia, Gamaleya Street 18, Moscow 123098, Russia

²I.M. Sechenov First Moscow State Medical University, Moscow 119991, Russia

Correspondence should be addressed to Olga L. Voronina; olv550@gmail.com

Received 24 February 2016; Accepted 8 November 2016

Academic Editor: Vassily Lyubetsky

Copyright © 2016 Olga L. Voronina et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Biofilm formation by *Burkholderia* spp. is a principal cause of lung chronic infections in cystic fibrosis patients. A “lacking biofilm production” (LBP) strain *B. contaminans* GIMC4587:Bct370-19 has been obtained by insertion modification of clinical strain with plasmid mutagenesis. It has an interrupted transcriptional response regulator (RR) gene. The focus of our investigation was a two-component signal transduction system determination, including this RR. *B. contaminans* clinical and LBP strains were analyzed by whole genome sequencing and bioinformatics resources. A four-component operon (BiofilmReg) has a key role in biofilm formation. The relative location (i.e., by being separated by another gene) of RR and histidine kinase genes is unique in BiofilmReg. Orthologs were found in other members of the Burkholderiales order. Phylogenetic analysis of strains containing BiofilmReg operons demonstrated evidence for earlier inheritance of a three-component operon. During further evolution one lineage acquired a fourth gene, whereas others lost the third component of the operon. Mutations in sensor domains have created biodiversity which is advantageous for adaptation to various ecological niches. Different species *Burkholderia* and *Achromobacter* strains all demonstrated similar BiofilmReg operon structure. Therefore, there may be an opportunity to develop a common drug which is effective for treating all these causative agents.

1. Introduction

The *Burkholderia cepacia* complex (*Bcc*) bacteria are opportunistic pathogens which cause nosocomial infections and are especially dangerous for cystic fibrosis (CF) patients. Analysis of *Bcc* strain diversity in Russian healthcare units and in CF patients demonstrated 5 species: *B. cenocepacia*, *B. multivorans*, *B. stabilis*, *B. contaminans*, and *B. vietnamiensis* [1]. Among these, *B. cenocepacia* was more abundant, and the Russian epidemic strain ST (sequence type) 709 belonged to this species. However, *B. contaminans* ST102 was also isolated from CF and non-CF patients [1]. Moreover, this strain is known to have an intercontinental spread across the world [2].

Presently, *Bcc* eradication is complex and, in most cases, impossible, which leads to chronic infections in the lungs

of CF patients. Biofilm formation is the principal reason for bacterial stability in CF patients' respiratory tracts.

The history of biofilm observation is long. But, in spite of studies of planktonic and aggregated forms of microbes that occurred hand in hand, the importance of the biofilm phenomena for medicine was first postulated by Hoiby and Axelsen only at the beginning of 1970s, based on observations of CF patients with chronic *Pseudomonas aeruginosa* lung infection [3]. Later, the first biofilm conference, in 1996, yielded better understanding of the significance of biofilm infection in medicine and marked the beginning of intensive microbial biofilm research. Many different approaches were used: conventional light microscopy; electron microscopy and confocal laser scanning microscopy for biofilm architecture and composition investigation; direct and accidental mutation of biofilm-forming strains; transcriptomic

analysis; and differential measurement of biochemical pathway activity and metabolite concentrations of planktonic and sessile cells [4]. Hence we now know that a large number of genes are involved in so complicated process of biofilm formation. For instance, comparison of high- and low-biofilm producing *B. pseudomallei* strains revealed 563 differentially regulated genes [5]. It should be noted that upregulated genes related to two-component signal transduction systems and a denitrification enzyme pathway [5].

A surprising result came from the work of Romanova et al. [6] on nondirectional insertion mutagenesis of high biofilm producer (HBP) clinical strain *B. contaminans* GIMC4509:Bct370, when just one of 1000 plasposon insertions had lost the ability to form biofilm. This LBP strain named *B. contaminans* GIMC4587:Bct370-19 had only one interrupted gene. DNA sequencing of a fragment adjacent to the insertion site identified it as the transcriptional regulator gene *ompR*, which is the part of the two-component signal transduction system (shortly two-component system, TCS). The TCS array consists of a protein histidine kinase (HK) and a response regulator (RR) protein. It is now known that TCSs mediate several different bacterial processes: chemotaxis, aerobic/anaerobic regulation, sporulation, and differentiation [7], as well as biofilm response [8].

The purpose of our investigation was the determination of this key in biofilm formation TCS, part of which is found transcriptional regulator. Detailed study of the BiofilmReg operon structure and evolution could have significant medical applications.

2. Materials and Methods

2.1. Bacterial Strains and Their Origins. All strains used came from the Gamaleya Institute Microbial Collection (GIMC): high biofilm producer (HBP) clinical strain *B. contaminans* GIMC4509:Bct370 (ST102, PubMLST id 1264) and lacking biofilm production (LBP) *B. contaminans* strain GIMC4587:Bct370-19. The LBP strain was obtained by insertion modification of clinical strain with plasposon pTnMod-RKm by Romanova et al. [6].

2.2. DNA Isolation and Genomics. Preparation of genomic DNA for the whole genome sequencing was performed as described [9]. Whole genome sequencing of *B. contaminans* strains was performed according to the manufacturer's (Roche) guidelines for the next generation sequencing (NGS). Two protocols were used for shotgun-sequencing library preparation: rapid library and pair-end library.

2.3. Data Acquisition and Processing. DNA sequence assembly into scaffolds was performed with 454 Sequencing System Software v.2.7 and v.3.0 (Roche). To aid in assembling individual chromosome we used data from reference strains: *B. lata* strain *Burkholderia* sp. 383; *B. contaminans* strain MS14; *B. ubonensis* strain MSMB22; *B. cenocepacia* strains: J2315, DDS 22E-1; *B. cepacia* strains: DDS 7H-2, ATCC 25416. The software Rapid Annotations Subsystems Technology

and SEED [10, 11] were used for annotating the genome of *B. contaminans* strains. BioProjects PRJNA349796 and PRJNA349797 were registered in GenBank with BioSample Accessions SAMN05933033 for GIMC4509:Bct370 and SAMN05933042 for GIMC4587:Bct370-19. Now the genomes are in the process of the chromosomes assembling.

2.4. Bioinformatic Analyses. Complementary protein description, prediction of domains, signal peptides, and protein cellular localization have been performed by NCBI BLAST [12], InterPro server [13, 14], TMHMM Server v. 2.0 [15], SignalP 4.1 Server [16], and PSORTb version 3.0.2. [17]. Promoter sequence prediction has been performed by BPROM (Prediction of Bacterial Promoters) [18, 19] and NNPP (Neural Network Promoter Prediction) [20] servers. Operon borders have been predicted with help of operon predictor: PTools04a (BioCyc Database Collection) [21]. Searches of the NCBI database for orthologs of operon components and gene was performed with the aid of the KEGG ORTHOLOGY (KO) Database [22, 23] and Biocyc Database [24]. Multiple alignments of nucleotide sequences were created in a MEGA 6.0 [25] environment using Multiple Sequence Alignment tools [26]. The numbers of nucleotide differences per site were counted as pairwise distances. Percent similarity and divergence coefficients have been determined by MegAlign 5.05 [25]. Amino acid sequence analysis tools in these same software packages were also used. Searches using NCBI BLAST have been performed for identification of significant sequences, containing sites of phosphorylation, intermolecular recognition, polypeptide, and DNA binding.

2.5. Phylogenetic Analysis. Phylogenetic analyses of polypeptide sequence data were performed in MEGA 6.0 [25]. Evolutionary history was inferred by using the Maximum Likelihood (ML) method based on the JTT matrix-based model [27]. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree for the heuristic search was obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model and then selecting the topology with superior log likelihood value. Trees were drawn to scale, with branch lengths measured as the number of substitutions per site. All positions containing gaps and missing data were eliminated.

Phylogenetic analysis of 16S rDNA nucleotide sequences was carried out in MEGA 6.0 [25]. The evolutionary history was inferred by using ML method based on the general time reversible model GTR+G [28], which was chosen as an optimal evolution distance model derived from Modeltest based on the Akaike information criterion [29]. Initial tree for the heuristic search was obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. A discrete Gamma distribution was used to model evolutionary rate differences among sites (6 categories (+G, parameter = 0.3082)). Bootstrap analyses were performed with 500 replicates.



FIGURE 1: BiofilmReg operon location on Burkholderia sp.383 chromosome 1 (GenBank: CP000151.1). A4580: 1555167...1555871, two-component transcriptional regulator, winged helix family; A4581: 1555868...1557229, peptidoglycan-binding LysM; A4582: 1557240...1559627, periplasmic sensor signal transduction histidine kinase; and A4583: 1559665...1560279, hypothetical protein (<http://www.ncbi.nlm.nih.gov/>). Blue triangle is in the position of the gene interruption by plasposon. Blue line marks the genes of operon.

3. Results and Discussion

3.1. Localization of Biofilm-Switch Response Regulator (RR) in *B. contaminans* Genomes. The clinical HBP strain *B. contaminans* GIMC4509:Bct370 and its modification, LBP strain *B. contaminans* GIMC4587:Bct370-19, were the objects of whole genome sequencing (WGS). Assembling the genomes in scaffolds allowed us to suggest a candidate position for the interruption position of the plasposon and then to find the neighbor genes of RR. Upstream of the insert site and on the same sense there were three genes with own promoters. The distance between the nearest outside gene and RR was 269 bp. This intergenic region included predicted promoter region, a transcription start site, and a 5' untranslated region (UTR). The promoter region was located at 164–213 bp upstream RR gene start codon, according to NNPP server. The positions of consensus –10 box and –35 box were detected at –192 bp and –212 bp, respectively, with help of BPROM server. The 5' UTR was surprisingly long for Prokaryotes: 173 bp. The suggestion of our prediction we found in Sass et al.'s experimental work [30]. Authors analyzed the RNA extracted from *B. cenocepacia* J2315 biofilm and revealed 187 CDS (coding sequence), which featured long 5' UTR of >150 nt, for transcriptional regulators, nucleotide binding, and membrane proteins. Among these 187 CDS was BCAL1443 (two-component regulatory system, response regulator protein) orthologous to our RR.

The three same sense downstream genes are genes of peptidoglycan-binding protein (PBP) with additional FecR domain, histidine kinase (HK), and uncharacterized protein (UnP) of DUF4136 superfamily (Figure 1). The next open reading frame (ORF) has been located on the antisense strand. The start codon of PBP gene was inside the RR coding region. The distance between the PBP and HK genes was 10 bp and 27 bp between HK and UnP genes. There was not any promoter downstream RR gene all the way to the first gene on the antisense strand according to promoter sequence prediction by BPROM and NNPP servers. So, basing our conclusion on the distances between the four adjacent genes in the same DNA strand and on the availability of a single transcriptional promoter we predict that their organization reflects a common transcriptional operon. Since the RR interruption by the plasposon had switched off the strain biofilm formation entirely, we had named the operon

“Biofilm Regulating” (shortly, BiofilmReg). The intact operon of HBP strain has been submitted to the NCBI GenBank database, with Accession Number KP288492. The LBP strain operon sequences have Accession Numbers KP288491 and KU252679.

As mentioned earlier, RR and HK are usually assumed to be a cognate pair, because they lie next to one another within the same operon [31]. In contrast, here the PBP gene is embedded between the RR and HK genes, so they are not “genomic neighbors” in Sheng et al.'s [32] interpretation, because the distance between them is more than 300 bp. Because some researchers who work on bioinformatic analysis of TCS in whole genomes might doubt our evidence that the HK gene from BiofilmReg operon is part of an operon with RR, we have done further investigations to show that these genes form natural functional units within a single operon.

3.2. Diversity of Two-Component Transcriptional Regulator (TCTR) Genes in *B. lata* Genome. Is this case of a coregulated gene inserted between RR and HK unique? To answer this question we searched for all two-component transcriptional regulator (TCTR) genes in the reference *B. lata* genome. We found 37 TCTR genes: 49% on chromosome 1 (3.69 Mb, NC_007510.1) and nearly equal numbers, 27 versus 24%, respectively, on chromosome 2 (3.59 Mb, INSDC NC_007511.1) and chromosome 3 (1.4 Mb, INSDC NC_007509.1) even though the second chromosome is twice as big as the third. Eleven TCTRs were found to be encoded by a single gene; the remaining 26 TCTRs were organized in operons. Among these 26, 22 operons were two-component types in accordance with the evidence that the average bacterial operon size is 2.2 genes [33]. The remaining four operons included more than two genes. The biggest contained eight genes whose products participated in phosphate transport. Another one was a three-component type located on the second chromosome. The last two operons had four-components. One of them included genes of the DUF4390 family uncharacterized protein and rRNA SAM-dependent methyltransferase as well as the RR and HK. The second one was the BiofilmReg operon. For the majority of these operons the gene adjacent to TCTR was HK. Only in the BiofilmReg operon was a PBP gene embedded between the RR and HK genes. Since gene organization in an operon is a means to coordinate expression functions [34] we next attempted to understand the possible functions of proteins encoded by BiofilmReg analyzing there domains.

3.3. Analysis of Proteins Domains in BiofilmReg Operon Components. First we analyzed the domains of the TCTR (Table S1, in Supplementary Material available online at <http://dx.doi.org/10.1155/2016/6560534>) in reference *B. lata* genome. According to the NCBI BLAST results, two conservative domains are present in TCTR: receiver and DNA binding. Together they form a multidomain polypeptide, having a Pfam classification [35]. Most of the TCTRs examined, about 70%, had a winged helix-turn-helix (wHTH) DNA binding domain (PF00486), 19% were the representatives of the LuxR family (PF00196), 5% were simple HTH.8

(PF02954) structures, and the last 5% belonged to the HTH-AraC (PF00165) family (Table S2). The most abundant group was subdivided into eight subgroups according to their multidomain characteristics. One of the common types, the CreB family, includes RR from the BiofilmReg operon; CreB is carbon source responsive response regulator that belongs to the CreBC two-component system. Investigation of this system in *E. coli* has demonstrated its participation in controlling genes involved in acetate [36] and ribose metabolism [37], in the maltose regulon [38], and in the pentose phosphate pathway [39] and genes which repair DNA damage associated with the replication fork [40]. Avison et al. [41] have named CreBC “the heart of metabolic regulation” in *E. coli* [41]. RR has localized in cytoplasm of bacterial cell.

The next component in the BiofilmReg operon that we examined is the gene for an uncharacterized conserved protein containing LysM and FecR domains. This is named according to COG4254 (clusters of orthologous groups) [42]. The InterPro resource classified this protein as an uncharacterized conserved protein UCP02964, LysM, PA4035. Orthologs of this gene were variously named: uncharacterized protein (UniProtKB U2H3R6), peptidase M23B (A0A0J6M8Q8), peptidoglycan-binding LysM (Q39H89), and FecR family protein (A0A088U8M6). The structure of LysM domain is known, and a function in peptidoglycan binding is predicted for it. It is found in a variety of enzymes involved in bacterial cell wall degradation [12]. The second domain is FecR, which is involved in regulation of iron dicitrate transport and is probably a sensor that recognizes iron dicitrate in the periplasm [12]. The InterPro service predicted for the protein product of the second component of BiofilmReg a signal peptide and the main part of the protein localized outside of the cytoplasm, which agrees with putative function for this domain.

Third gene in BiofilmReg operon is a gene for a histidine kinase or periplasmic sensor signal transducer histidine kinase. This HK is a multidomain protein. The first domain starting from the N-terminus is transmembrane; the second is a CHASE2 domain (pfam05226), which is an extracellular sensory domain. Environmental factors that are recognized by CHASE2 domains are not known at this time [12]. The next three HK structural elements are transmembrane domains. The subsequent PAS domain is a second sensor domain, which is not present in all HK types [43]. This adaptable domain can monitor changes in light, redox potential, oxygen, or small ligands, depending on their associated cofactor [7]. PAS domain is localized in cytoplasm. The next two domains have the same localization. These are (1) dimerization and phosphotransfer and (2) catalytic and ATP-binding. All together these last domains form multidomain. According to COG classification (COG3852) the HK from BiofilmReg operon is nitrogen specific, having multidomain NtrB [12].

The fourth and final component of the operon is the gene for an uncharacterized protein with a DUF4136 domain. This domain has been found in bacterial lipoproteins [12]. According to InterPro this polypeptide has a signal peptide and the main part of the protein has external localization.

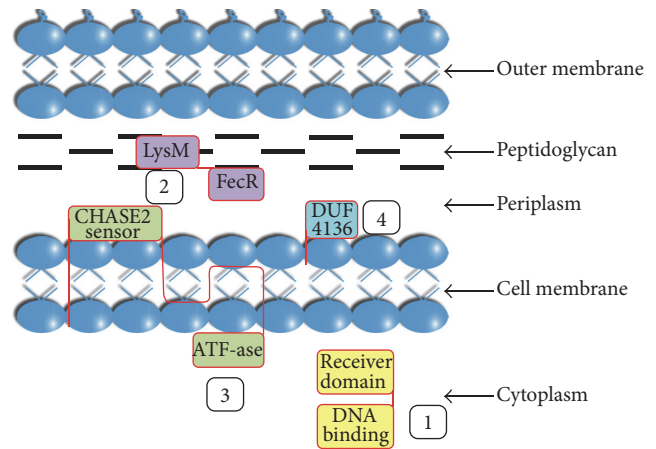


FIGURE 2: Components of BiofilmReg operon topology in bacterial cell predicted by InterPro Service (<http://www.ebi.ac.uk/interpro/>). The components are as follows: (1) two-component transcriptional regulator; (2) uncharacterized conserved protein, containing LysM and FecR domains; (3) periplasmic sensor signal transduction histidine kinase; and (4) uncharacterized DUF4136 superfamily protein.

3.4. A Model of BiofilmReg Protein Localization. In summary, we present a proposal for colocalization of the four described proteins in bacterial cells (Figure 2).

The periplasm contains two proteins and the signaling domain of HK. One is binding to the rigid exoskeleton (peptidoglycan), which determines cell shape [44], the second is bound to the lipids of the outer or inner membranes. The PBP and HK units sense different signals, which can be transmitted to RR in the cytoplasm and alter target gene expression.

The interruption of this operon by plasmid pTnMod-RKm insertion destroyed all four genes' transcription. Only short fragment of RR gene (148 bp) rests before plasmid sequence. Promoter, detected by BPROM and NNPP servers in the end of plasmid, is divided by the second part of RR gene (564 bp) from the next ORF and so cannot be active. As a result Romanova et al. detected the absence of biofilm formation in vitro by modified strain [6].

3.5. Searching of Orthologs of BiofilmReg Operon Components. Do such operons occur frequently in other known bacteria? We searched for orthologs of BiofilmReg operon components to answer this question. First we analyzed Gram-negative bacteria of classes Beta- and Gammaproteobacteria, which are often recorded among opportunistic microorganisms that cause nosocomial infections. A cohort of 45 genomes belonging to 21 genera was examined. A result was considered positive if at least two adjacent components of the operon were detected together in the panel of genomes we searched (Table 1).

Among Gammaproteobacteria, only two *Pseudomonas* strains, *P. aeruginosa* PAO1 and *P. fluorescens* PCL1751, had a couple of orthologous genes. These appear to be an exception. In the class Betaproteobacteria, only the Burkholderiales order had genera containing orthologs of the BiofilmReg

TABLE 1: Representatives of Betaproteobacteria and Gammaproteobacteria, which were checked for presence of at least two adjacent components of operon.

| Class | Present operon components | Order/family | Absent operon components |
|---------------------|---|----------------------------------|--|
| Betaproteobacteria | <i>Burkholderia contaminans</i> GIMC4509:Bct370, <i>Burkholderia lata</i> str. <i>Burkholderia</i> sp. 383, <i>Burkholderia contaminans</i> MS14, <i>Burkholderia dolosa</i> AU0158, <i>Burkholderia multivorans</i> ATCC 17616, <i>Burkholderia vietnamiensis</i> G4, <i>Burkholderia cenocepacia</i> J2315, HI2424, <i>Lautropia mirabilis</i> ATCC 51599, <i>Pandoraea thiooxydans</i> DSM 25325 | Burkholderiales Burkholderiaceae | <i>Pandoraea pnomenusa</i> 3 kgm |
| | <i>Alcaligenes faecalis</i> subsp. <i>faecalis</i> NCIB 8687, <i>Achromobacter xylosoxidans</i> NH44784-1996, NCTC10807, MN001, <i>Achromobacter ruhlandii</i> SCCH3: Ach s33-1365, ST36, <i>Achromobacter insuavis</i> AXX-A | Burkholderiales Alcaligenaceae | — |
| | <i>Bordetella bronchiseptica</i> MOI49, <i>Bordetella pertussis</i> 18323 <i>Ralstonia pickettii</i> 12], <i>Ralstonia pickettii</i> DTP0602 | Burkholderiales Ralstoniaceae | — |
| | — | Burkholderiales Comamonadaceae | <i>Variovorax paradoxus</i> SI10, <i>Acidovorax avenae avenae</i> ATCC 19860 |
| | — | Burkholderiales Sutterellaceae | <i>Sutterella parvirubra</i> YTT |
| | — | Neisseriales | <i>Neisseria meningitidis</i> WUE 2594 |
| | <i>Pseudomonas aeruginosa</i> PAO1, <i>Pseudomonas fluorescens</i> PCL1751 | Pseudomonadales | <i>Acinetobacter haemolyticus</i> ATCC 19194, <i>Pseudomonas fluorescens</i> A506, SBW25, BRIP34879, F113, Pf0-1, BBc6R8 |
| | — | Enterobacteriales | <i>Yersinia intermedia</i> ATCC 29909, <i>Yersinia pestis</i> KIM D27, KIM10+, <i>Escherichia coli</i> K-12 substr. MG1655, <i>Shigella flexneri</i> 2a str. 2457T, 2a str. 301, <i>Salmonella enterica enterica</i> serovar Dublin str. CT_02021853 |
| | — | Xanthomonadales | <i>Xanthomonas campestris</i> pv. <i>campestris</i> str. ATCC 33913 |
| | — | Legionellales | <i>Legionella pneumophila pneumophila</i> Hextuple-2q, Hextuple-3a, Thunder Bay |
| Gammaproteobacteria | — | Vibrionales | <i>Vibrio cholerae</i> CIP1030(3) |

operon genes. Various species of the *Burkholderia* genus included fully sized operons in their genomes (Tables 1 and 2). However, other genera in the Burkholderiaceae family had individual species with orthologs of the operon.

Thus, *Pandoraea thiooxydans* contained a couple of the target genes (Table 2), but there were no orthologous genes in *P. pnomenusa* (CP006900.2). Three genera of Alcaligenaceae and a single species, *Ralstonia pickettii*, from the Ralstoniaceae family had orthologs of operon genes. Representatives of two other families of the order Burkholderiales—Comamonadaceae and Sutterellaceae—had no genomes with orthologs. Overall 21 genomes with orthologous genes for BiofilmReg operon have been identified and analyzed (Table 2). Among them there is the Russian epidemic strain *Achromobacter ruhlandii* ST36 (GenBank Accession Number CP017433.1) [45], whose operon is submitted in GenBank with Accession Number KU252680. Almost all *Burkholderia* genomes have an identical operon structure: RR, PBP, HK, and UnP. However, in *L. mirabilis* (Burkholderiaceae) and *A. faecalis* (Alcaligenaceae), genome UnP was substituted by glutamyl-tRNA reductase and AraC family transcriptional regulator, respectively. Representatives of *Achromobacter*, *Bordetella*, and *Ralstonia* genera had three-component operons without UnP. Finally, the operon of *P. thiooxydans* DSM 25325 was the most divergent in Burkholderiales order and included only two genes: PBP and HK.

In one gammaproteobacterium a related operon was found in the *P. aeruginosa* PAO1 genome. This consisted of just PBP with a truncated LysM domain and an intact FecR domain, plus HK. Despite the alteration in PBP it was classified by NCBI BLAST analysis as COG4254 too [12].

Given the interesting distribution of these operons we now asked: What is their evolutionary history? To reconstruct the original operon structure in their common ancestral bacterium we analyzed the phylogeny of the listed microorganisms with help of 16S rDNA gene sequences as a base of biosystematics [46].

3.6. Phylogeny of Burkholderiales Representatives and *P. aeruginosa* Based on 16S rDNA Gene Sequences. A Maximum Likelihood 16S rDNA gene tree (Figure 3) has been created for 50 sequences, which included all identified 16S rDNA gene copies of 21 representatives of the families Burkholderiaceae, Alcaligenaceae, Ralstoniaceae, and Pseudomonadaceae (Table 2). It should be noted that 16S rDNA gene copies of some genomes have differences in the sequence, so the number of 16S rDNA gene copies increases in more than two times the number of genomes in the analysis.

The phylogenetic tree revealed two main clusters of the Burkholderiales representatives, corresponding to the Alcaligenaceae and Burkholderiaceae families. Representatives of Ralstoniaceae were embedded inside the Burkholderiaceae cluster as a distinct group. The Alcaligenaceae family cluster (Bootstrap Index, BI, 100%) included *Achromobacter*, *Bordetella*, and *Alcaligenes* species, and the Burkholderiaceae cluster (BI 98%) included *Burkholderia*, *Pandoraea*, *Lautropia*, and *Ralstonia* species. It should be noted that representatives of each genus formed separate clades inside the two major clusters. The *P. aeruginosa* PAO1 operon is the most divergent

2-component operon with an altered PBP and a normal HK. In contrast to *P. aeruginosa* PAO1 (Pseudomonadaceae, Gammaproteobacteria), which is treated as the outgroup, all Burkholderiales genomes contain orthologous operons with at least three components: RR, HK, and PBP being thus consistent with them having a common origin. The three-component operon structure was observed in representatives of two families Ralstoniaceae (*R. pickettii*) and Alcaligenaceae (*A. xylosoxidans*, *A. ruhlandii*, *A. insuavis*, *B. bronchiseptica*, and *B. pertussis*). And four-component operon was detected predominantly in representatives of Burkholderiaceae (*B. contaminans*, *B. lata*, *B. dolosa*, *B. multivorans*, *B. vietnamiensis*, *B. cenocepacia*, and *L. mirabilis*) and only in one *A. faecalis* strain of Alcaligenaceae. So it is clear that the three-component operon (RR, PBP, and HK) represents the ancestral state for the major clusters.

It is interesting to trace the evolution of whole operon and its components across the different taxa. For example, in the genome *Bordetella pertussis* 18323 the damaged HK gene cannot code catalytic domain and remains present in the operon only as a pseudogene. In genus *Ralstonia* we detected operon only in some *R. pickettii* strains.

This species differs in chromosome number from others in the genus *Ralstonia*. Here, *R. solanacearum* has only one chromosome, *R. mannitolilytica* has two, and *R. pickettii* has three chromosomes, as do most representatives of the Burkholderia genus which are dangerous for CF patients and for patients of the department of reanimation and intensive therapy as nosocomial infection. One strain, *R. pickettii* 12J, had a conventional version of the three-component operon (Table 2) located on chromosome I. However, in the genome of *R. pickettii* DTP0602, three-component operon, has been found on chromosome II, indicating the possible translocation of a full-sized operon. The PBP structure of this strain has provided additional support for this suggestion. The PDP sequence had an additional fragment at its C-terminal end, which was identified as COG4733, phage-related protein, tail component [12].



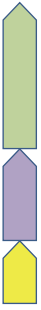





In the *P. thiooxydans* DSM 25325 genome, the two operon genes (PBP and HK) lie together on the same sense strand as usual, but the orthologous RR gene is duplicated 3' and set in the reverse direction on the complementary DNA strand. This arrangement suggests complex recombination and translocation events.









The *Burkholderia* species, *L. mirabilis* (Burkholderiaceae), and *A. faecalis* (Alcaligenaceae) have probably each independently gained an extra (fourth) operon component during their evolution. In contrast to the general similarity of RR, HK, and PBP between these operons, the extra component of the *A. faecalis* operon is an AraC family transcriptional regulator, while the fourth component of the *L. mirabilis* operon is a glutamyl-tRNA reductase [12]. The extra component of all other *Burkholderia* operons belongs to the DUF4136 superfamily, whose function is still unknown. The various functions of the fourth components may indicate that these operons have been recruited in different metabolic pathways, probably involving different signal perception and transduction functions depending on bacterial lifestyle. The presence of these components in operons from different phylogenetic

TABLE 2: Main domain of the proteins, coding by BiofilmReg operon genes.

| N | 1 | 2 | 3 | 4 | Strain name (GenBank: Accession Number) | Operons schematic presentation |
|----|--|---|--|--|---|--------------------------------|
| 1 | 1126...1830 (KP288492.1)/ PRK11083 (CreB) | 1827...3188 (KP288492.1)/ LysM, FecR | 1...2388 (KU252679)/ PAS9, NtrB | 2426...3040 (KU252679)/ DUF4136 | Burkholderia contaminans strain GIMC4509-Bct370 (KP288492.1; KU252679) | |
| 2 | BCEP18194_RS13000/ PRK11083 (CreB) | BCEP18194_RS13005/ LysM, FecR | BCEP18194_RS13010/ PAS9, NtrB | BCEP18194_RS13015/ DUF4136 | Burkholderia lata strain Burkholderia sp. 383 chromosome 1, complete sequence (NC_007510.1) | |
| 3 | NL30_RS08520/ PRK11083 (CreB) | NL30_RS08515/ LysM, FecR | NL30_RS08510/ PAS9, NtrB | NL30_RS08505/ DUF4136 | Burkholderia contaminans strain MS14 chromosome 1, complete sequence (NZ_CP009743) | |
| 4 | AK34_RS21645/ PRK11083 (CreB) | AK34_RS21640/ LysM, FecR | AK34_RS21635/ PAS9, BaeS | AK34_RS21630/ DUF4136 | Burkholderia dolosa AU0158 chromosome 1, complete sequence (NZ_CP009795) | |
| 5 | BMULJ_RS06770/ PRK11083 (CreB) | BMULJ_RS06775/ LysM, FecR | BMULJ_RS06780/ PAS9, BaeS | BMULJ_RS06785/ DUF4136 | Burkholderia multivorans ATCC 17616 DNA, complete genome, chromosome 1 (NC_010804.1) | |
| 6 | BCEP1808_RS07040/ PRK11083 (CreB) | BCEP1808_RS07045/ LysM, FecR | BCEP1808_RS07050/ PAS9, NtrB | BCEP1808_RS07055/ DUF4136 | Burkholderia vietnamiensis G4 chromosome 1, complete sequence (NC_009256.1) | |
| 7 | QU43_RS43720/ PRK11083 (CreB) | QU43_RS43725/ LysM, FecR | QU43_RS43730/ PAS9, NtrB | QU43_RS43735/ DUF4136 | Burkholderia cenocepacia J2315 chromosome 1, complete genome (NC_011000.1) | |
| 8 | BCEN2424_RS07115/ PRK11083 (CreB) | BCEN2424_RS07120/ LysM, FecR | BCEN2424_RS07125/ PAS9, NtrB | BCEN2424_RS07130/ DUF4136 | Burkholderia cenocepacia HI2424 chromosome 1, complete sequence (NC_008542.1) | |
| 9 | HMPREF0551_RS12390/ <u>PRK10643 (BaeR)</u> | HMPREF0551_RS12385/ LysM, FecR | HMPREF0551_RS12380/ <u>PAS absent, BaeS</u> | HMPREF0551_RS12375/ <u>glutamyl-tRNA reductase</u> | Lautropia mirabilis ATCC 51599 genomic scaffold SCAFFOLD1, whole genome shotgun sequence (NZ_GL636062.1) | |
| 10 | QWA_RS04640/ PRK11083 (CreB) | QWA_RS04635/ LysM, FecR | QWA_RS04630/ <u>PAS absent, BaeS</u> | QWA_RS04625/ <u>AraC family transcriptional regulator</u> | Alcaligenes faecalis subsp. faecalis NCIB 8687 Contig.3, whole genome shotgun sequence (NZ_AKMR01000003.1) | |
| 11 | NH44784_RS11280/ <u>PRK10336 (QseB)</u> | NH44784_RS11275/ LysM, FecR | NH44784_RS11270/ <u>PAS4, BaeS</u> | — | Achromobacter xylosoxidans NH44784-1996 complete genome (NC_021285.1) | |
| 12 | ERS451415_06153/ PRK11083 (CreB) | ERS451415_06154/ LysM, FecR | ERS451415_06155/ PAS9, BaeS | — | Achromobacter xylosoxidans NCTC10807 (LN831029.1) | |
| 13 | Axylo_5268/ PRK11083 (CreB) | Axylo_5269/ LysM, FecR | Axylo_5270/ PAS9, BaeS | — | Achromobacter xylosoxidans strain MN001, complete genome (CP012046.1) | |

TABLE 2: Continued.

| N | 1 | 2 | 3 | 4 | Strain name (GenBank: Accession Number) | Operons schematic presentation |
|----|--|--|---|---|---|---|
| 14 | Complement (1269...3539)/ PRK11083 (CreB) | Complement (3601...4710)/ LysM, FecR | Complement (4744...5460)/ PAS9, BaeS | — | <i>Achromobacter ruhlandii</i> SCCH3: Ach 33-1365, ST36 (KU252680) |  |
| 15 | AXXA_RS10090/ <u>PRK10336 (QseB)</u> | AXXA_RS10095/ LysM, FecR | AXXA_RS10100/ <u>PAS4, BaeS</u> | — | <i>Achromobacter insuavis</i> AXX-A genomic scaffold scaffold000003, whole genome shotgun sequence (NZ_GL982453.1) |  |
| 16 | BN115_RS00125/ PRK11083 (CreB) | BN115_RS00120/ LysM, FecR | BN115_RS00115/ PAS9, BaeS | — | <i>Bordetella bronchiseptica</i> MO149 complete genome (NC_018829.1) |  |
| 17 | BN118_RS00125/ PRK11083 (CreB) | BN118_RS00120/ LysM, FecR | BN118_RS00115/ PAS9, BaeS , <u>no catalytic domain</u> | — | <i>Bordetella pertussis</i> 18323 complete genome (NC_018518.1) |  |
| 18 | RPIC_RS04635/ PRK11083 (CreB) | RPIC_RS04640/ LysM, FecR | RPIC_RS04645/ <u>PAS4, BaeS</u> | — | <i>Ralstonia pickettii</i> 12J chromosome 1, complete sequence (NC_010682.1) |  |
| 19 | N234_RS31485/ PRK11083 (CreB) | LysM, FecR, COG4733, phage-related protein, tail component ABW99_RS09030/ LysM, FecR | N234_RS31475/ <u>PAS4, BaeS</u> | — | <i>Ralstonia pickettii</i> DTP0602 chromosome 2, complete sequence (NC_022514.1) |  |
| 20 | — | ABW99_RS09035/ <u>PAS4, BaeS</u> | — | — | <i>Pandoraea thiooxydans</i> DSM 25325, complete genome (NZ_CP011568.1) |  |
| 21 | — | PA4035/ Lys M truncated, FecR | PA4036/ <u>PAS4, BaeS</u> | — | <i>Pseudomonas aeruginosa</i> PAOI chromosome, complete genome (NC_002516.2) |  |

-  (1) Two-component transcriptional regulator, winged helix family
 (2) Uncharacterized conserved protein, containing LysM and FecR domains
 (2') Uncharacterized conserved protein with truncated LysM and intact FecR domains
 (3) Periplasmic sensor signal transduction histidine kinase
 (4) Uncharacterized DUF4136 superfamily protein
 (4') Glutamyl-tRNA reductase
 (4'') AraC family transcriptional regulator
 Pseudogene

* Differences in domain/operon organization or localization are underlined.

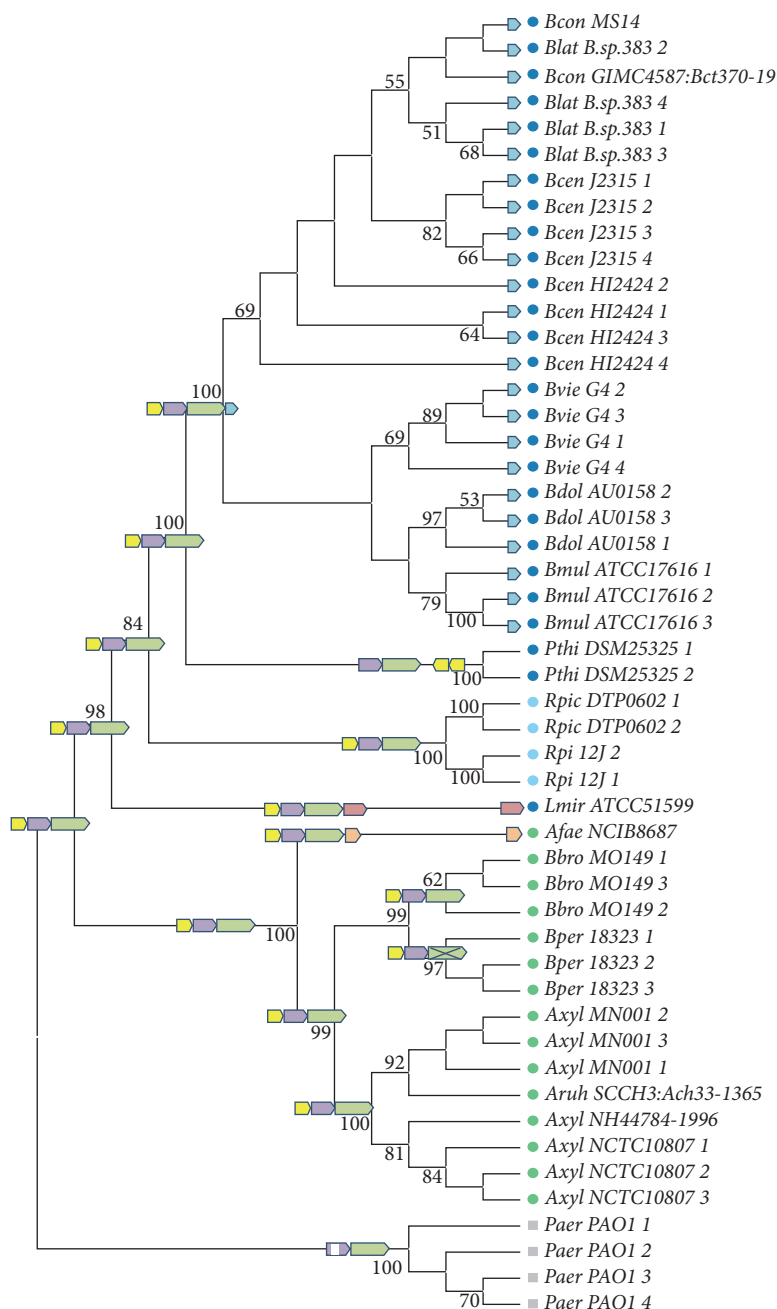


FIGURE 3: ML phylogenetic tree of Burkholderiaceae representatives and *P. aeruginosa* based on 16S rDNA gene sequences. Dark blue: Burkholderiaceae, blue: Ralstoniaceae, green: Alcaligenaceae, and grey: *Pseudomonas aeruginosa*. Operons schematic representation is given on the branch nodes according to Table 1 symbols.

lineages among Burkholderiaceae and Alcaligenaceae representatives also supports the view that these were gained by them rather than lost from all other phylogenetic lineages.

If the conventional version of the three-component operon has an ancient origin, then we can ask: How much genetic distinction has accumulated in operon genes during evolution along different phylogenetic lineages? To answer this question we analyzed the polypeptides encoded by a selection of these operons.

3.7. Analysis of the Sequence Diversity in Individual Operon Components. Our first proteins comparison was done at the level of domains identified by NCBI BLAST. Most RR proteins belonged to the CreB subfamily (Table 2) except for three strains: *L. mirabilis* with an RR of the BasR subfamily (PRK10643) and *A. xylosoxidans* NH44784-1996 plus *A. insuavis* AXX-A with RRs of the QseB subfamily (PRK10336). The second component, PBP, includes two domains LysM and FecR in most cases. Characteristics of the *L. mirabilis*

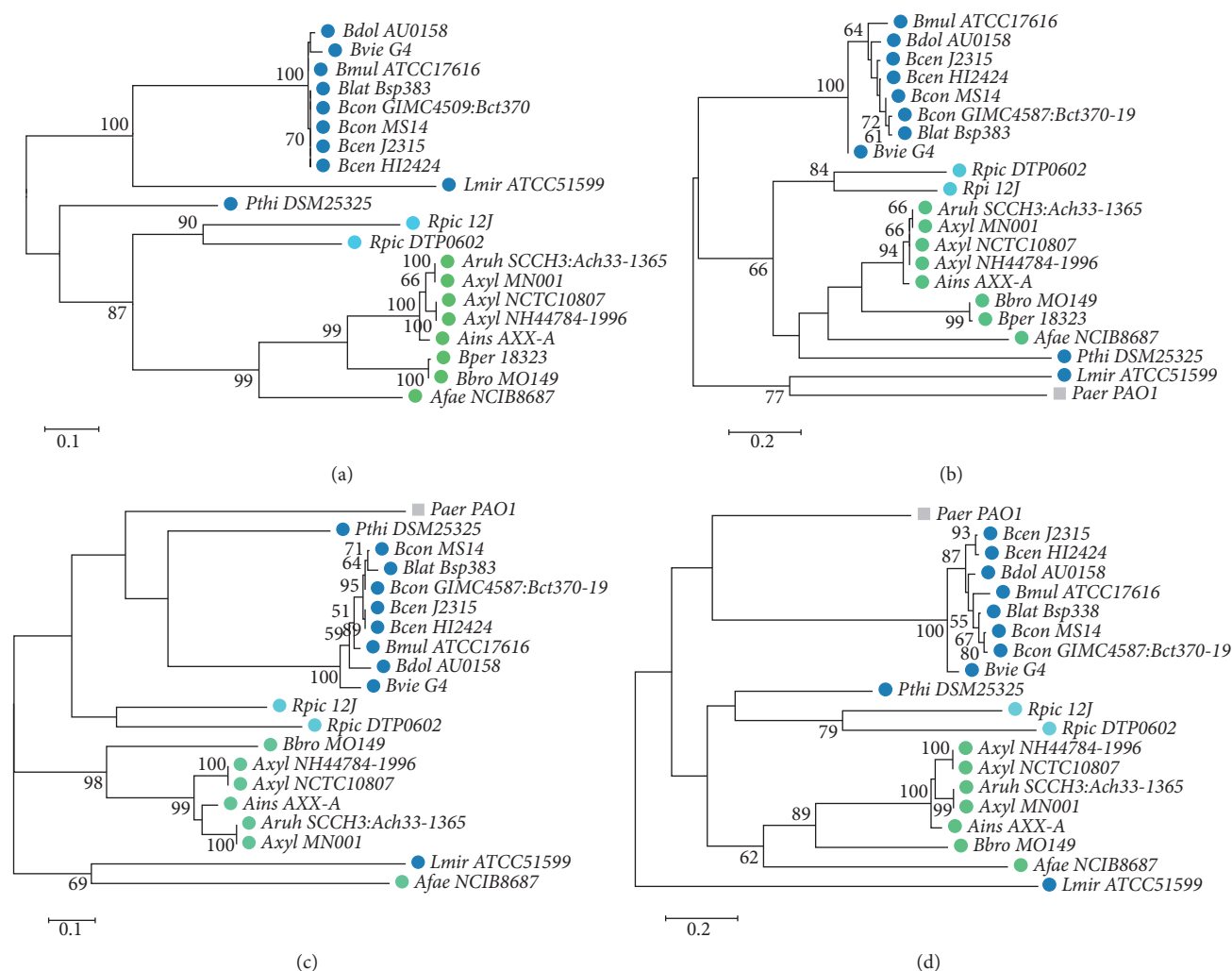


FIGURE 4: ML phylogenetic trees based on aligned amino acid characters of (a) response regulator (REC signal receiver domain and trans_reg_C effector domain), (b) uncharacterized conserved protein, containing LysM and FecR domains, (c) histidine kinase PAS + HisKA + HATPase_c domains, and (d) histidine kinase CHASE2 domain. Dark blue: Burkholderiaceae, blue: Ralstoniaceae, green: Alcaligenaceae, and grey: *Pseudomonas aeruginosa*. 228 aligned characters of response regulator (REC signal receiver domain + trans_reg_C effector) (a), 213 and 199 aligned characters of histidine kinase (PAS + HisKA + HATPase_c and CHASE2, resp.) (b, c), and 113 aligned characters of peptidoglycan-binding protein (LysM and FecR domain) (d) were taken for phylogeny reconstruction.

and *P. aeruginosa* PBP components were discussed earlier in Section 3.6. The third component, HK, has an extracellular sensory CHASE2 domain (pfam05226) in all strains. However, the cytoplasmic domains are variable. The second sensor domain, PAS9, was replaced by PAS4 in six strains from different phylogenetic lineages and absent in *L. mirabilis* and in *A. faecalis*. The next two domains, forming multidomain of NtrB subfamily (COG3852) in BiofilmReg operon HK of LBP strain, belong in most cases to the BaeS subfamily (COG0642) even in genus *Burkholderia* (Table 2).

The second level of proteins comparison consisted of evaluating phylogenetic relatedness for individual operon components based on the most conservative protein regions. Four resulting alignments are shown in Figure 4. The four resulting ML phylogenetic trees had slightly different topologies.

In general, the trees included two clusters: a *Burkholderia* species cluster and an *Achromobacter* + *Bordetella* cluster. The positions of *P. aeruginosa*, *L. mirabilis*, *A. faecalis*, *R. pickettii*, and *P. thiooxydans* were variable and sometimes unconventional. For instance, *R. pickettii* strains and *P. thiooxydans* DSM 25325 were not merged with Burkholderiaceae species, in contrast to the more generally accepted phylogeny of these taxa. Most RRs exhibited greater similarity than HKs: 38.2% to 100% compared with 31.8–100% and 24.9–100% percent similarity for PAS + HisKA + HATPase_c and CHASE2 domains of HKs, respectively. It can be elucidated by variable specificity of the HK sensor domains.

The *Burkholderia* genus was more representative in our analysis, so we compared proteins similarities inside this genus alone. Most variable among them were the CHASE domains of HK (differences: 1.2–13%), PBP (3.7–24.1%),

and UnP (2.4–23.4%) sequences. All these protein domains are localized in the periplasm and have first contacted with signal molecules. Five strains (*Burkholderia* sp. 383, MS14, G4, HI2424, and ATCC 17616) in our cohort were environment from different ecological niches and three strains (GIMC4509:Bct370, AU0158, and J2315) were host-associated: CF or non-CF patients. We suggest that the variability sequences as revealed here may reflect special adaptive characteristics of the *Burkholderia* strains.

3.8. Horizontal Gene Transfer versus Foreign DNA Contamination? The *Burkholderia* species themselves and other Burkholderiales representatives are the primary soil-dwelling bacteria successfully specialized to different ecological niche, including host-associations. The presence of *P. aeruginosa* has been observed in all these niches, which could explain the acquisition of the BiofilmReg operon by an ancestral *P. aeruginosa* strain. Moreover, *Burkholderia*, *Achromobacter*, *Ralstonia*, and *Pseudomonas* species were previously included together in one genus, *Pseudomonas*. Only the advent of molecular-genetic methods allowed microbiological systematics to split this huge assemblage.

Surprisingly, we found an orthologs of full-size BiofilmReg operon in single genome from Gram-positive bacterium, the actinobacterium *Mumia flava* strain MUSC 201. This is a new genus in the family Nocardioideaceae, which was first approved in 2014 [47]. This strain was originally isolated from mangrove soil in Malaysia. Because horizontal gene transfer (HGT) is a well-known contributor to gene exchange between bacteria, Archaea, and Eukarya, we considered that this might be an example of this process. Orthologs of the BiofilmReg operon have been found in contig 65 of whole shotgun genome *Mumia flava* MUSC 201 (JTDJ01000001–JTDJ01000923) with similarity for RR, PBP, HK, and UnP genes 100.0%, 95.1%, 97.0%, and 96.6%, respectively. However, some observations were highly enigmatic. For example, the *Mumia flava* genome was unexpectedly big: 16.4 Mb, in contrast to the few other Nocardioideaceae genomes that have been assembled into chromosomes, with sizes from 3.1 to 7.6 Mb. Second, contig 65 was very similar in sequence to the reference *B. lata* genome, not only within the borders of the operon but along a 180 kb stretch. Third, in *M. flava* contigs 1, 12, 134, 150, and 229 we found sequences similar to those of *Burkholderia cenocepacia* J2315, *Burkholderia contaminans* MS14, and *Ralstonia pickettii* 12D, including their 16S rDNA sequences. So we obtained the evidence for foreign DNA pollution in *M. flava* strain MUSC 201 genome and made sure exclusively chromosome assembled genomes are verified material for gene analysis.

This means only some Gram-negative bacteria have orthologs of BiofilmReg operon.

4. Conclusion

The four-component operon of *Burkholderia contaminans* strain GIMC4509:Bct370, named BiofilmReg, was intriguing by biofilm switching ability and structure organization. It was shown to be unique with respect to the relative locations RR and HK in its operon. Exact orthologs of this operon were

found only in the Burkholderiales order of Gram-negative bacteria and not in two *Pseudomonas* strains. Phylogenetic analysis base of 16S rDNA gene sequences and in accordance with the operon structure demonstrated the evidence of three-component operon inheritance from an ancestral bacterium. During evolution, one lineage acquired a fourth gene and others lost the third component. Mutations, especially in sensor domains, helped to increase biodiversity and allow for adaptation to various ecological niches. So now we can observe *Burkholderia*, *Achromobacter*, and *Ralstonia* species as emerging pathogens. This is a result of shift from living free in a natural habitat to adoption of a host-associated pathogen lifestyle [48]. Multiple antibiotic resistance and biofilm formation help these strains avoid therapeutic drugs. Because *Burkholderia* and *Achromobacter* strains from different species all demonstrated a similar operon structure, there is an opportunity to develop a common drug for all these causative agents.

Abbreviations

| | |
|-----------------|---|
| Bcc: | <i>Burkholderia cepacia</i> complex |
| CF: | Cystic fibrosis |
| GIMC: | The Gamaleya Institute Microbial Collection |
| ST: | Sequence type |
| HBP: | High biofilm producer |
| LBP: | Lacking biofilm production |
| NGS: | Next generation sequencing |
| WGS: | Whole genome sequencing |
| NCBI: | The National Center for Biotechnology Information |
| UTR: | Untranslated region |
| CDS: | Coding sequence |
| TCS: | Two-component signal transduction system (shortly two-component system) |
| RR: | Response regulator |
| HK: | Histidine kinase |
| PBP: | Peptidoglycan-binding protein |
| UnP: | Uncharacterized protein |
| ML: | Maximum Likelihood |
| ORF: | Open reading frame |
| BiofilmReg: | Biofilm Regulating |
| TCTR: | Two-component transcriptional regulator |
| wHTH: | Winged helix-turn-helix |
| COG: | Clusters of orthologous groups |
| BI: | Bootstrap Index |
| Blat_B.sp.383: | <i>Burkholderia lata</i> strain <i>Burkholderia</i> sp. 383 |
| Bcon_MS14: | <i>Burkholderia contaminans</i> strain MS14 |
| Bdol_AU0158: | <i>Burkholderia dolosa</i> AU0158 |
| Bmul_ATCC17616: | <i>Burkholderia multivorans</i> ATCC 17616 |
| Bvie_G4: | <i>Burkholderia vietnamiensis</i> G4 |
| Bcen_J2315: | <i>Burkholderia cenocepacia</i> J2315 |
| Bcen_HI2424: | <i>Burkholderia cenocepacia</i> HI2424 |

| | |
|------------------------|--|
| Lmir_ATCC51599: | <i>Lautropia mirabilis</i> ATCC 51599 |
| Afae_NCIB8687: | <i>Alcaligenes faecalis</i> subsp. <i>faecalis</i> NCIB 8687 |
| Axyl_NH44784-1996: | <i>Achromobacter xylosoxidans</i> NH44784-1996 |
| Axyl_NCTC10807: | <i>Achromobacter xylosoxidans</i> NCTC10807 |
| Axyl_MN001: | <i>Achromobacter xylosoxidans</i> MN001 |
| Aruh_SCCH3:Ach33-1365: | <i>Achromobacter ruhlandii</i> SCCH3:Ach33-1365 |
| Ains_XXX-A: | <i>Achromobacter insuavis</i> XXX-A |
| Bbro_MO149: | <i>Bordetella bronchiseptica</i> MO149 |
| Bper_18323: | <i>Bordetella pertussis</i> 18323 |
| Rpic_12J: | <i>Ralstonia pickettii</i> 12J |
| Rpic_DTP0602: | <i>Ralstonia pickettii</i> DTP0602 |
| Pthi_DSM25325: | <i>Pandoraea thiooxydans</i> DSM 25325 |
| Paer_PAO1: | <i>Pseudomonas aeruginosa</i> PAO1. |

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by Grant 14-04-00116 of the Russian Foundation for Basic Research “Two-Component Signal Transduction System OmpR/EnvZ is a Target for Directed Search Antibacterial Drugs against Biofilm Infection.”

References

- [1] O. L. Voronina, M. S. Kunda, N. N. Ryzhova et al., “The variability of the order burkholderiales representatives in the healthcare units,” *BioMed Research International*, vol. 2015, Article ID 680210, 9 pages, 2015.
- [2] *The Main Site PubMLST hosted at The Department of Zoology, University of Oxford*, Oxford, UK, 2016, <http://pubmlst.org/>.
- [3] N. Hoiby and N. H. Axelsen, “Identification and quantitation of precipitins against *Pseudomonas aeruginosa* in patients with cystic fibrosis by means of crossed immunoelectrophoresis with intermediate gel,” *Acta Pathologica et Microbiologica Scandinavica. Section B: Microbiology*, vol. 81, pp. 298–308, 1973.
- [4] N. Hoiby, “A personal history of research on microbial biofilms and biofilm infections,” *Pathogens and Disease*, vol. 70, no. 3, pp. 205–211, 2014.
- [5] C.-Y. Chin, Y. Hara, A.-K. Ghazali et al., “Global transcriptional analysis of *Burkholderia pseudomallei* high and low biofilm producers reveals insights into biofilm production and virulence,” *BMC Genomics*, vol. 16, no. 1, article 471, 2015.
- [6] Y. M. Romanova, I. G. Tiganova, I. A. Khmel et al., “Biofilms of burkholderia cepacia: characteristics of mutants with modified formation capacity,” *Molecular Genetics, Microbiology and Virology*, vol. 26, no. 3, pp. 93–101, 2011.
- [7] A. M. Stock, V. L. Robinson, and P. N. Goudreau, “Two-component signal transduction,” *Annual Review of Biochemistry*, vol. 69, pp. 183–215, 2000.
- [8] E. Karatan and P. Watnick, “Signals, regulatory networks, and materials that build and break bacterial biofilms,” *Microbiology and Molecular Biology Reviews*, vol. 73, no. 2, pp. 310–347, 2009.
- [9] K. Wilson, *UNIT 2.4 Preparation of Genomic DNA from Bacteria*, Current Protocols in Molecular Biology, Wiley Online Library, Hoboken, NJ, USA, 2001.
- [10] R. K. Aziz, D. Bartels, A. Best et al., “The RAST Server: rapid annotations using subsystems technology,” *BMC Genomics*, vol. 9, article 75, 2008.
- [11] R. Overbeek, T. Begley, R. M. Butler et al., “The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes,” *Nucleic Acids Research*, vol. 33, no. 17, pp. 5691–5702, 2005.
- [12] A. Marchler-Bauer, M. K. Derbyshire, N. R. Gonzales et al., “CDD: NCBI’s conserved domain database,” *Nucleic Acids Research*, vol. 43, no. 1, pp. D222–D226, 2015.
- [13] A. Mitchell, H. Y. Chang, L. Daugherty et al., “The InterPro protein families database: the classification resource after 15 years,” *Nucleic Acids Research*, vol. 43, no. 1, pp. D213–D221, 2015.
- [14] P. Jones, D. Binns, H.-Y. Chang et al., “InterProScan 5: genome-scale protein function classification,” *Bioinformatics*, vol. 30, no. 9, pp. 1236–1240, 2014.
- [15] Center for Biological Sequence Analysis Server, Hosted in Technical University of Denmark, <http://www.cbs.dtu.dk/services/TMHMM/>.
- [16] T. N. Petersen, S. Brunak, G. von Heijne, and H. Nielsen, “SignalP 4.0: discriminating signal peptides from transmembrane regions,” *Nature Methods*, vol. 8, no. 10, pp. 785–786, 2011.
- [17] N. Y. Yu, J. R. Wagner, M. R. Laird et al., “PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes,” *Bioinformatics*, vol. 26, no. 13, pp. 1608–1615, 2010.
- [18] V. Solovyev and A. Salamov, “Automatic annotation of microbial genomes and metagenomic sequences,” in *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*, R. W. Li, Ed., pp. 61–78, Nova Science, 2011.
- [19] Site Softberry, Inc, Mount Kisco, NY, USA, <http://www.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb>.
- [20] Site UC Berkeley/University of Hohenheim, Berkeley, UC; Stuttgart, Germany, http://www.fruitfly.org/seq_tools/promoter.html.
- [21] P. R. Romero and P. D. Karp, “Using functional and organizational information to improve genome-wide computational prediction of transcription units on pathway-genome databases,” *Bioinformatics*, vol. 20, no. 5, pp. 709–717, 2004.
- [22] KEGG Orthology (KO) Database, Hosted in Kanehisa Laboratories, Kyoto University and University of Tokyo, Japanese, <http://www.kegg.jp/kegg/ko.html>.
- [23] M. Kanehisa, *Post-Genome Informatics*, Oxford University Press, 2000.
- [24] R. Caspi, T. Altman, R. Billington et al., “The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases,” *Nucleic Acids Research*, vol. 42, no. 1, pp. D459–D471, 2014.
- [25] K. Tamura, G. Stecher, D. Peterson, A. Filipski, and S. Kumar, “MEGA6: molecular evolutionary genetics analysis version 6.0,” *Molecular Biology and Evolution*, vol. 30, no. 12, pp. 2725–2729, 2013.
- [26] The Main Site EMBL-EBI, European Bioinformatics Institute, 2016, <http://www.ebi.ac.uk/Tools/msa/clustalw2>.

- [27] D. T. Jones, W. R. Taylor, and J. M. Thornton, "The rapid generation of mutation data matrices from protein sequences," *Computer Applications in the Biosciences*, vol. 8, no. 3, pp. 275–282, 1992.
- [28] M. Nei and S. Kumar, *Molecular Evolution and Phylogenetics*, Oxford University Press, New York, NY, USA, 2000.
- [29] D. Posada and K. A. Crandall, "MODELTEST: testing the model of DNA substitution," *Bioinformatics*, vol. 14, no. 9, pp. 817–818, 1998.
- [30] A. M. Sass, H. Van Acker, K. U. Förstner et al., "Genome-wide transcription start site profiling in biofilm-grown *Burkholderia cenocepacia* J2315," *BMC Genomics*, vol. 16, article 775, 2015.
- [31] L. Burger and E. van Nimwegen, "Accurate prediction of protein-protein interactions from sequence alignments using a Bayesian method," *Molecular Systems Biology*, vol. 4, article 165, 2008.
- [32] X. Sheng, M. Huvel, J. W. Pinney, and M. P. Stumpf, "Evolutionary characteristics of bacterial two-component systems," in *Evolutionary Systems Biology*, vol. 751 of *Advances in Experimental Medicine and Biology*, chapter 6, pp. 121–137, Springer, Berlin, Germany, 2012.
- [33] M. Madan Babu, Ed., *Bacterial Gene Regulation and Transcriptional Networks*, Caister Academic Press, Poole, UK, 2013.
- [34] F. Jacob and J. Monod, "Genetic regulatory mechanisms in the synthesis of proteins," *Journal of Molecular Biology*, vol. 3, pp. 318–356, 1961.
- [35] R. D. Finn, A. Bateman, J. Clements et al., "Pfam: the protein families database," *Nucleic Acids Research*, vol. 42, no. 1, pp. D222–D230, 2014.
- [36] H. Kakuda, K. Hosono, K. Shiroishi, and S. Ichihara, "Identification and characterization of the *ackA* (acetate kinase A)-*pta* (phosphotransacetylase) operon and complementation analysis of acetate utilization by an *ackA-pta* deletion mutant of *Escherichia coli*," *Journal of Biochemistry*, vol. 116, no. 4, pp. 916–922, 1994.
- [37] C. A. Dunn, S. F. O'Handley, D. N. Frick, and M. J. Bessman, "Studies on the ADP-ribose pyrophosphatase subfamily of the Nudix hydrolases and tentative identification of *trgB*, a gene associated with tellurite resistance," *Journal of Biological Chemistry*, vol. 274, no. 45, pp. 32318–32324, 1999.
- [38] E. Richet, "On the role of the multiple regulatory elements involved in the activation of the *Escherichia coli* *malE* promoter," *Journal of Molecular Biology*, vol. 264, no. 5, pp. 852–862, 1996.
- [39] G. A. Sprenger, "Genetics of pentose-phosphate pathway enzymes of *Escherichia coli* K-12," *Archives of Microbiology*, vol. 164, no. 5, pp. 324–330, 1995.
- [40] C. J. Saveson and S. T. Lovett, "Tandem repeat recombination induced by replication fork defects in *Escherichia coli* requires a novel factor, *RadC*," *Genetics*, vol. 152, no. 1, pp. 5–13, 1999.
- [41] M. B. Avison, R. E. Horton, T. R. Walsh, and P. M. Bennett, "*Escherichia coli* *CreBC* is a global regulator of gene expression that responds to growth in minimal media," *The Journal of Biological Chemistry*, vol. 276, no. 29, pp. 26955–26961, 2001.
- [42] R. L. Tatusov, E. V. Koonin, and D. J. Lipman, "A genomic perspective on protein families," *Science*, vol. 278, no. 5338, pp. 631–637, 1997.
- [43] A. E. Bem, N. Velikova, M. T. Pellicer, P. V. Baarlen, A. Marina, and J. M. Wells, "Bacterial histidine kinases as novel antibacterial drug targets," *ACS Chemical Biology*, vol. 10, no. 1, pp. 213–224, 2015.
- [44] T. J. Silhavy, D. Kahne, and S. Walker, "The bacterial cell envelope," *Cold Spring Harbor Perspectives in Biology*, vol. 2, no. 5, Article ID a000414, 2010.
- [45] O. L. Voronina, M. S. Kunda, N. N. Ryzhova et al., "Diversity and hazard of respiratory infection of *Achromobacter* spp. in cystic fibrosis patients," *Pulmonology*, vol. 25, no. 4, pp. 389–402, 2015 (Russian).
- [46] C. R. Woese, "Bacterial evolution," *Microbiological Reviews*, vol. 51, no. 2, pp. 221–271, 1987.
- [47] L.-H. Lee, N. Zainal, A.-S. Azman, N.-S. Ab Mutalib, K. Hong, and K.-G. Chan, "*Mumia flava* gen. nov., sp. nov., an actinobacterium of the family *Nocardioidaceae*," *International Journal of Systematic and Evolutionary Microbiology*, vol. 64, no. 5, pp. 1461–1467, 2014.
- [48] B. Zhu, M. Ibrahim, Z. Cui et al., "Multi-omics analysis of niche specificity provides new insights into ecological adaptation in bacteria," *The ISME Journal*, vol. 10, pp. 2072–2075, 2016.

Research Article

Resistance of Permafrost and Modern *Acinetobacter lwoffii* Strains to Heavy Metals and Arsenic Revealed by Genome Analysis

Sofia Mindlin,¹ Anatolii Petrenko,¹ Anton Kurakov,¹ Alexey Beletsky,²
Andrey Mardanov,² and Mayya Petrova¹

¹Institute of Molecular Genetics, Russian Academy of Sciences, Kurchatov Sq. 2, Moscow 123182, Russia

²Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, Leninsky Ave. 33, Bld. 2, Moscow 119071, Russia

Correspondence should be addressed to Mayya Petrova; petrova@img.ras.ru

Received 20 April 2016; Revised 14 July 2016; Accepted 7 September 2016

Academic Editor: Peter F. Stadler

Copyright © 2016 Sofia Mindlin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We performed whole-genome sequencing of five permafrost strains of *Acinetobacter lwoffii* (frozen for 15–3000 thousand years) and analyzed their resistance genes found in plasmids and chromosomes. Four strains contained multiple plasmids (8–12), which varied significantly in size (from 4,135 to 287,630 bp) and genetic structure; the fifth strain contained only two plasmids. All large plasmids and some medium-size and small plasmids contained genes encoding resistance to various heavy metals, including mercury, cobalt, zinc, cadmium, copper, chromium, and arsenic compounds. Most resistance genes found in the ancient strains of *A. lwoffii* had their closely related counterparts in modern clinical *A. lwoffii* strains that were also located on plasmids. The vast majority of the chromosomal resistance determinants did not possess complete sets of the resistance genes or contained truncated genes. Comparative analysis of various *A. lwoffii* and of *A. baumannii* strains discovered a number of differences between them: (i) chromosome sizes in *A. baumannii* exceeded those in *A. lwoffii* by about 20%; (ii) on the contrary, the number of plasmids in *A. lwoffii* and their total size were much higher than those in *A. baumannii*; (iii) heavy metal resistance genes in the environmental *A. lwoffii* strains surpassed those in *A. baumannii* strains in the number and diversity and were predominantly located on plasmids. Possible reasons for these differences are discussed.

1. Introduction

The presence of plasmids, extrachromosomal self-replicating circular DNA molecules, in most bacterial species is connected to their key role in rapid adaptation to changing environmental conditions without altering the gene content of the bacterial chromosome. Typically, in addition to few essential genes a plasmid harbors genes coding for ecologically important properties such as variable metabolic processes, nitrogen fixation, resistance to antibiotics and heavy metals, and pathogen virulence.

In recent years, with the revolution in sequencing technologies the number of sequenced bacterial plasmids increased dramatically and reached about 4600 [1], but only a small number of these plasmids have been analyzed in detail.

In particular, not enough attention was paid to such important issues as evolution and dynamics of plasmids in natural bacterial populations, despite a number of works that were devoted to this research [2–7]. Bacteria belonging to the *Acinetobacter* genus are a convenient model for such studies since these are ubiquitous bacteria that play an important role in different ecological niches including soil, water, and association with animals and, at the same time, some of them are common human pathogens. For instance, Fondi et al. [8] have undertaken a study of evolutionary relationships between 29 *Acinetobacter* plasmids using the information on sequenced plasmids. It should be noted that the authors included in the list of analyzed sequences eight plasmids containing deficient mercury resistance transposons described in our previous work [9]. The authors demonstrated the important

TABLE 1: Characteristics of permafrost *Acinetobacter lwoffii* strains.

| Strain | Age of sediment* | Sampling locality |
|---------|------------------|---|
| ED23-35 | 20–40 K | Kolyma Lowland, Bank of river Homus-Yuryiah, Late-Pleistocene-Ice-Complex |
| ED45-23 | 20–40 K | Kolyma Lowland, Bank of river Homus-Yuryiah, Late-Pleistocene-Ice-Complex |
| ED9-5a | 15–30 K | Kolyma Lowland, Bank of river Homus-Yuryiah, Late-Pleistocene-Ice-Complex |
| EK30A | 1,6–1,8 M | Kolyma Lowland, Bank of river Konkovaya, Early Pleistocene |
| VS15 | 2-3 M | Kolyma Lowland, Bank of river Grand Chukochia, Late-Pleistocene-Ice-Complex |

* Age of subsoil permafrost sediments (corresponding to the period of freezing before the present) in years: K = 10^3 years; M = 10^6 years.

role of rearrangements between different plasmids and concluded that transposases and selective pressure for mercury resistance have played a pivotal role in plasmid evolution in *Acinetobacter*. Unfortunately, only small regions (less than 10%) of the mercury resistance *Acinetobacter* plasmids were sequenced at that time and this fact was not taken into account. In addition, the selected plasmids did not represent adequately the diversity of plasmids hosted by different strains belonging to the *Acinetobacter* genus.

Recently, we could continue our studies on the molecular structure and phylogenetic relations of *Acinetobacter*'s plasmids using next-generation sequencing methods. For our studies we have chosen five ancient strains belonging to *Acinetobacter lwoffii*, guided by the following considerations: (i) in comparison with plasmids of *A. baumannii*, plasmids of *A. lwoffii* have been studied very little; (ii) in the last years, *A. lwoffii* strains were included in the list of dangerous pathogens to humans.

Surprisingly, analysis of genomic sequences demonstrated that all the ancient strains isolated from uncontaminated permafrost sediments contained plasmids with genes of resistance to salts of heavy metals and arsenic. We hypothesized that not only mercury but also other heavy metals have played and continue to play an important role in the evolution of *Acinetobacter*. This prompted us to study the occurrence and abundance of such genes in the genomes of *A. lwoffii* isolates. Here we describe the results of analysis of the structure and distribution of heavy metals and arsenic resistance genes found in the genomes of ancient as well as of modern strains of *A. lwoffii*.

2. Materials and Methods

2.1. Bacterial Strains and Growth Conditions. The ancient *Acinetobacter lwoffii* strains used in this study (Table 1) were previously isolated from permafrost sediments collected from different regions of Kolyma Lowland [9, 10]. All strains were grown in lysogeny broth (LB) medium or solidified LB medium (LA) [11] at 30°C.

2.2. Whole-Genome Sequencing and Assembly of Plasmids. The genomes of the ancient *A. lwoffii* strains were sequenced with a Roche GS FLX genome sequencer (Roche, Switzerland) using the Titanium protocol to prepare shotgun genomic libraries. Approximately 40-fold sequence coverage

was achieved for each genome. The GS FLX reads were assembled into contigs using the GS De Novo Assembler (Roche); among them, we identified those contigs that contained genes for mobilization and/or replication of plasmids. PCR was used to close the gaps between assembled plasmid contigs and to confirm their circular structure.

2.3. Bioinformatics Analysis. For the assembly and analysis of plasmid genomes from ancient strains the program UGENE (<http://unipro.ru/>) was used. Similarity searches were performed using BLAST [12] and REBASE [13]. Open reading frames (ORFs) were searched using ORF Finder and BLAST software at NCBI. Conserved domains and motifs were identified using the NCBI Conserved Domain Database (CDD) [14] and the Pfam database [15]. Contigs from the whole-genome shotgun sequences of clinical *A. lwoffii* strains were considered chromosomal if their size exceeded 300 kb and if they contained housekeeping genes. Relatively small (less than 200 kb) contigs were attributed to plasmids or their parts if they contained determinants typical for plasmids and/or if they contained no housekeeping genes.

2.4. Standard DNA Manipulations. Standard protocols were used for DNA isolation and agarose gel electrophoresis [11]. PCR was performed with a Mastercycler (Eppendorf) using Taq DNA polymerase and Long PCR Enzyme Mix (for amplification fragments more than 6 kbp) with supplied buffers (Thermoscientific), dNTP mixture (Thermoscientific), and appropriate primers. Amplified DNA fragments were sequenced using conventional Sanger method (Applied Biosystems 3730 Genetic Analyzer) at the Interinstitute Genome Center (Moscow).

2.5. Determination of the MICs for Heavy Metals and Arsenic Salts' Resistance. The level of resistance was determined by the agar dilution method [16]. The MICs of heavy metals and arsenic salts for the ancient *A. lwoffii* strains were determined. Bacterial strains were grown in LB broth at 30°C with shaking for 3 h and diluted 100-fold with fresh broth. 5 μ L of the bacterial suspension (about 5×10^6 cells per mL) was plated with a bacteriological loop onto LA (to determine the MICs of As₃, As₅, Hg, and Cr) or Adams [11] (to determine the MICs of Cu, Co, Cd, and Zn) plates supplemented with heavy metals and arsenic salts. The used salts and their concentrations (mM) tested were as follows: AsNaO₂ (As₃): 5, 10, 20, and 30; Na₂HAsO₄ \times 7H₂O (As₅): 20, 40, and 60; K₂Cr₂O₇

TABLE 2: Characteristics of ancient plasmids carrying heavy metal resistance determinants and corresponding MICs.

| Strain | Plasmid [GenBank accession number] | Size of plasmid (bp) | Resistance determinants | MICs* |
|---------|------------------------------------|----------------------|-------------------------|---|
| ED23-35 | pALWED1.1 [KX426227] | 287630 | Hg, Chr, Co/Zn/Cd, Ni | Hg(>0.06) , As3(<5), As5(20), Cr(1.8) , Co(0.1) , Cd(0.025) , Zn(6.4) , Ni(2.7) , Cu(<0.9) |
| | pALWED1.3 [KX426228] | 16071 | Chr | |
| ED45-23 | pALWED2.1 [KX426229] | 190039 | Hg, Ars, Cu | Hg(0.06) , As3(20) , As5(40) , Cr(<0.3), Co(<0.05), Cd(<0.012), Zn(<0.2), Ni(<0.45), Cu(0.9) |
| ED9-5a | pALWED3.1 [KX528687] | 138028 | Hg, Ars, Co/Zn/Cd, Cu | Hg(>0.06) , As3(30) , As5(60) , Cr(1.2) , Co(<0.05), Cd(0.025) , Zn(3.2) , Ni(1.8), Cu(3.6) |
| | pALWED3.5 [KX426230] | 16568 | Chr | |
| VS15 | pALWVS1.1 [KX426232] | 134767 | Co/Zn/Cd, Cu | Hg(<0.015), As3(<5), As5(<20), Cr(<0.3), Co(0.05) , Cd(0.1) , Zn(3.2) , Ni(0.9), Cu(3.6) |
| EK30A | pALWEK1.1 [KX528688] | 209983 | Co/Zn/Cd, Cu | Hg(<0.015), As3(<5), As5(<20), Cr(1.2) , Co(0.05) , Cd(0.025) , Zn(<0.2), Ni(0.45), Cu(3.6) |
| | pALWEK1.5 [KX426231] | 8227 | Chr | |

*The minimal inhibitory concentrations (mM) of each salt are presented in parentheses; resistance due to the presence of the relevant genes marked in bold.

(Cr): 0.3, 0.6, 0.9, 1.2, 1.5, and 1.8; $\text{CuSO}_4 \times 5\text{H}_2\text{O}$ (Cu): 0.9, 1.8, 2.7, 3.6, and 4.0; HgCl_2 (Hg): 0.015, 0.03, and 0.06; $\text{CdCl}_2 \times \text{H}_2\text{O}$ (Cd): 0.01, 0.025, and 0.05; $\text{CoCl}_2 \times 6\text{H}_2\text{O}$ (Co): 0.01, 0.05, and 0.1; $\text{ZnSO}_4 \times 7\text{H}_2\text{O}$ (Zn): 0.2, 0.4, 0.8, 1.6, 3.2, and 6.4; $\text{NiSO}_4 \times 7\text{H}_2\text{O}$ (Ni): 0.45, 0.9, 1.8, and 2.7. The plates were incubated at 30°C for 24–30 h and visually inspected.

3. Results and Discussion

3.1. Whole-Genome Sequencing of Five Ancient Permafrost Strains of *A. lwoffii* and Identification of the Resistance Plasmids. During analysis of the whole-genome shotgun contigs of five ancient permafrost strains of *A. lwoffii* (Table 1), we paid the main attention to assembly and identification of plasmids. In total, to date we have identified 35 plasmids. It was found that 4 of 5 bacterial strains contained 8–12 plasmids and one strain carried 2 plasmids. The plasmids varied considerably in size (from 4,135 to 287,630 bp) and structural features. In our previous work, we identified and described in detail one of the small plasmids, pALWED1.8 with streptomycin/spectinomycin gene *aadA27* [17]. In the present work, we focused on the study of resistance genes to heavy metals and arsenic salts found in all the plasmids. Therefore, we first identified the plasmids harboring genes encoding resistance to salts of mercury (*mer* operons), arsenic (*ars* operons), chromium, copper (*cop* locus), and cobalt/zinc/cadmium (*czc* determinants).

Although all strains were isolated from pristine permafrost sediments, each of them contained one or two plasmids carrying different combinations of resistance genes to heavy metals and arsenic salts (Table 2).

We hypothesize that the abundance and diversity of the resistance genes may be associated with the environmental

habitat of the ancient strains. We decided to test the functional activity of the detected resistance genes, to analyze the structure of the resistance determinants found in ancient plasmids, and to check the presence of the resistance genes and their localization in genomes of modern strains of *A. lwoffii*.

3.2. Functional Activity of Resistance Genes. To test the activity of resistance determinants identified in the ancient plasmids the levels of resistance to salts of mercury, arsenic, chromium, copper, cobalt, zinc, cadmium, and nickel of five permafrost *A. lwoffii* strains were examined. For this purpose, MICs for different salts were determined by the agar dilution method (see Section 2). We carried out five independent experiments. All strains containing resistance genes were able to grow at higher concentrations of corresponding toxic compounds compared with strains not containing resistance determinants (Table 2). Thus, it can be concluded that at the high probability all detected resistance genes are functionally active. However, *czc* determinants in some strains provided resistance to salts of not all three metals (cobalt/zinc/cadmium), but only two of them. For example, the strain ED9-5a was sensitive to cobalt and EK30A to zinc salts. These facts can be explained by the existence of a complex transcriptional regulation of different components of the *czc* system described in [18].

3.3. Structure of Resistance Determinants Revealed on the Plasmids from Ancient *A. lwoffii* Strains

3.3.1. Mer Operons. The *mer* operons of Gram-negative bacteria contain a metalloregulator gene (*merR*) and three structural genes: genes that encode a transport system that delivers the toxic mercuric ions into the cells (*merTP*) and a gene that encodes an intracellular enzyme, mercuric reductase (*merA*),

TABLE 3: Heavy metal resistance operons identified in permafrost strains of *A. lwoffii*.

| Operon | Operon genes | Plasmid | Closest relative (AC) | Nucleotide sequence identity (%) |
|-------------|---|--|--|----------------------------------|
| <i>mer</i> | <i>merR</i> , <i>T</i> , <i>P</i> , <i>C</i> , <i>A</i> , <i>D</i> , <i>E</i> , (<i>B</i> [*]) | pALWED1.1 pALWED2.1 pALWED3.1 | <i>A. lwoffii</i> , pKLH202 (AJ486857) | 99% |
| <i>ars</i> | <i>trxB</i> , <i>arsH</i> , <i>B</i> , <i>C</i> , <i>R</i> , <i>C</i> | pALWED2.1 pALWED3.1 | <i>A. johnsonii</i> XBB1 (CP010350) | 78% |
| <i>cop</i> | <i>copD</i> , <i>C</i> , <i>F1</i> , <i>S</i> , <i>R</i> , <i>A</i> , <i>B</i> | pALWED2.1 pALWED3.1 pALWVS1.1 pALWEK1.1 | <i>A. baumannii</i> AB0057 (CP001182) | 76% |
| <i>czc1</i> | <i>czcC</i> , <i>B</i> , <i>A</i> , <i>D</i> | pALWVS1.1 pALWEK1.1 | <i>A. guillouiae</i> NBRC 110550 (AP014630) | 74% |
| <i>czc2</i> | <i>czcC</i> , <i>B</i> , <i>A</i> , <i>D</i> , <i>rcnR</i> , <i>nreB</i> | pALWED1.1 | <i>A. johnsonii</i> XBB1 (CP010350) | 82% |
| <i>czc3</i> | <i>czcC</i> , <i>B</i> , <i>A</i> , <i>D</i> | pALWED3.1 | <i>Moraxella osloensis</i> CCUG 350 (CP014234) | 97% |
| <i>chr1</i> | <i>chrA</i> - <i>chrB</i> | pALWED1.3 pALWED3.5 pALWEK1.5 | <i>A. sp</i> M131, pM131-6 (JX101643) | 99% |
| <i>chr2</i> | <i>chrA</i> - <i>chrB</i> | pALWED1.1 | <i>A. sp</i> M131, pM131-6 (JX101643) | 71% |

* *merB* gene is present only in the plasmid pALWED1.1.

which converts mercuric ions into less toxic metallic mercury [19]. Thus, the minimum set of essential mercury resistance genes is *merRTPA*; other genes (*merC*, *merF*, *merD*, *merE*, *orf2*, and *orfY*) occur in various combinations in the *mer* operons of Gram-negative bacteria and are not essential accessory genes. Mercury resistance operons and transposons of *Acinetobacter* strains have been the subject of our previous studies. In particular, *mer* operon of permafrost *A. lwoffii* strain ED23-35 studied in this work was described in detail in 2004 [9]. It was shown that in addition to the standard set of *mer* genes (*merRTPCADE*) characteristic to *Acinetobacter* strains it contains gene *merB* encoding organomercurial lyase. In this study, we found *mer* operons in two more strains resistant to mercury, ED45-23 and ED9-5a (Table 3). *Mer* operons of these strains were almost identical to each other and unlike strain ED23-35 did not contain the *merB* gene (Figure 1).

3.3.2. Chromium Resistance Determinants. Genetic analysis of chromate resistant bacteria *Pseudomonas aeruginosa* [20] and *Alcaligenes eutrophus* [21] has shown that main genes encoding the chromium resistance are genes *chrA*, encoding the chromate transporter protein, and *chrB*, encoding chromate resistance protein, probably performing the functions of a regulator. A hydrophobic membrane protein, *ChrA*, catalyzes the active efflux of chromate (chromate efflux pump) from the cytoplasm and chromate or dichromate induces expression of the *chrA* gene through the action of *chrB* [22]. It should be noted that, besides *chrA* and *chrB*, some accessory genes have been found in transposons and plasmids of chromate resistance bacteria.

Previously, chromium resistant bacteria were isolated from various groups of the genus *Acinetobacter*, in particular from *A. lwoffii* [23]. However, the resistance determinants in these bacteria were not described.

We revealed the *chrA* and *chrB* genes in three permafrost strains, ED23-35, ED9-5a, and EK30A (Tables 2 and 3). In all these strains chromium resistance genes were located on small plasmids pALWED1.3, pALWED3.5, and pALWEK1.5, respectively, and were almost identical (99% identity at the nucleotide sequence level) on their molecular structure. At the same time, these plasmids differed by their length as well as by the set and structure of essential genes. In particular, plasmid pALWED3.5 contained a *mobA* gene while in two others this gene was absent.

Interestingly, strain ED23-35 contained one more pair of the *chrA*-*chrB* genes. They were located on the largest plasmid pALWED1.1 (Figure 1) and differed significantly from the chromium resistance genes located on the small plasmids in their structure (identity about 60–70% at the nucleotide sequence level). Moreover, we could not reveal genes closely related to this variant of the *chrAB* genes in public databases.

Thus it can be concluded that chromium resistance genes *chrAB* are widely spread in environmental *A. lwoffii* strains. Because almost identical genes are present on different plasmids harbored by different *A. lwoffii* strains it can be proposed that they can spread via horizontal gene transfer.

3.3.3. Copper Resistance Determinants. Copper is needed for cell metabolism but is toxic at elevated concentrations. This challenge has led to the evolution of complex mechanisms of copper resistance in bacteria. Now a number of different

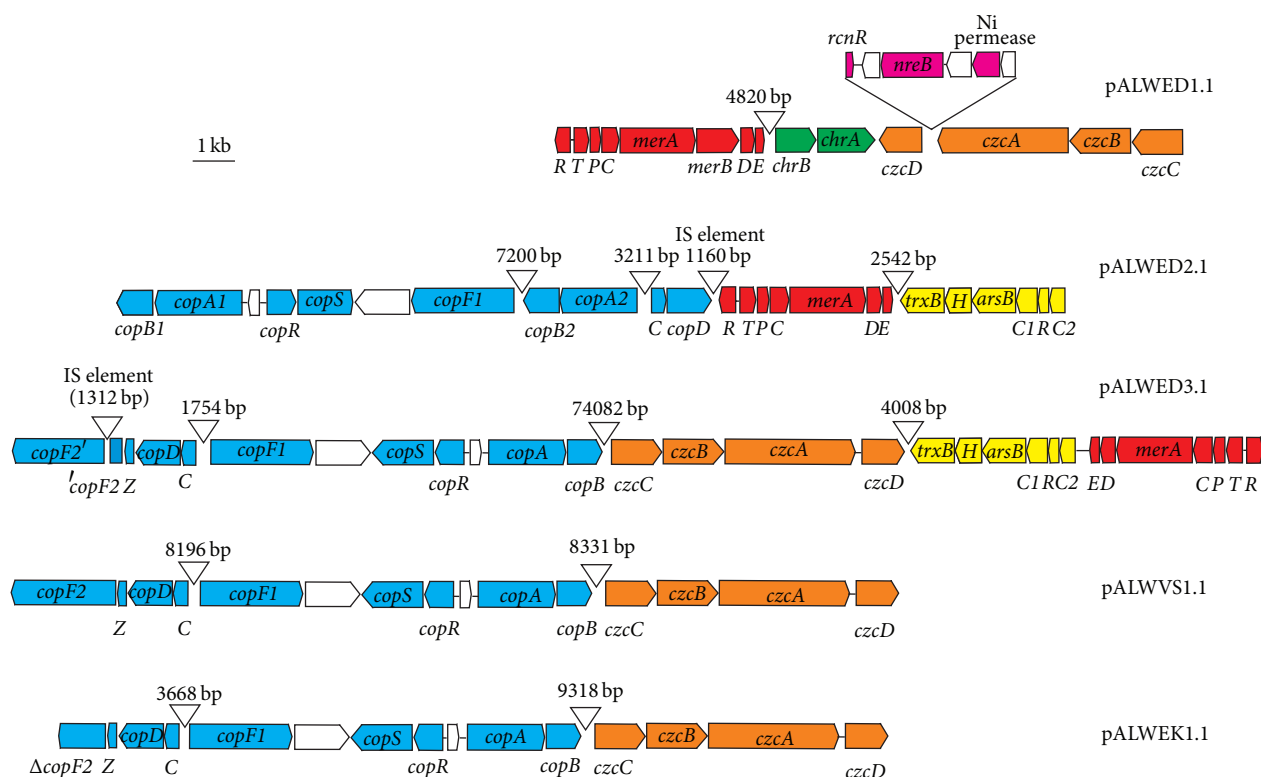


FIGURE 1: Genetic structure of resistance regions detected in the five large permafrost multiresistance plasmids (scheme). The location and polarity of genes are shown with arrows; genes belonging to *mer* operon are marked in red; *ars* operon in yellow; chromium resistance genes in green; *cop* genes in blue; *czc* in orange; nickel resistance genes in magenta; and genes with unknown functions and foreign genes in white. Numbers show the size of the sequences located between different resistance determinants and genes.

systems that allow bacterial cells to survive in the presence of high concentration of copper have been identified in different taxa of bacteria [24–28].

We found copper resistance determinants on plasmids from four permafrost *A. lwoffii* strains. They were located on the large plasmids (pALWED2.1, pALWED3.1, pALWVS1.1, and pALWEK1.1) (Table 3). We studied genetic structure of the resistance determinants harbored by pALWVS1.1 and revealed that they are characterized by unique features. We found that they are closely related to the copper resistance determinants of *Acinetobacter baumannii* AB0057 (CP001182.1) and *Acinetobacter johnsonii* XBB1 (CP010350.1) which are located on their chromosomes and form a single cluster. In addition to genes *copABCD* this cluster contains genes encoding copper-exporting ATPase (*copF*), heavy metal sensor kinase (*copS*), and transcriptional activation protein (*copR*). An unusual characteristic of the copper resistance determinants of *Acinetobacter* strains is the location of the *copAB* and *copCD* genes at a considerable distance from each other, on both sides of other genes involved in copper resistance (Figure 1); in other bacterial species these four genes are linked together and are parts of the same operon [26, 27, 29].

It should be noted that the cluster of copper resistance genes revealed in pALWVS1.1 in comparison with the above-mentioned chromosomal clusters contained an extra fragment about 5000 bp inserted into the region located between

the *copC* and *copF* genes (Figure 1). In this fragment we detected genes encoding two transposases and integrase. The same clusters of copper resistance genes were detected in pALWED3.1 and pALWEK1.1.

The genes of copper resistance found in pALWED2.1 are closely related to those described above but form a more complex mosaic structure as a result of the insertion of a large DNA fragment (about 12 kb). This insert contained (i) an extra copy of the *copA-copB* genes (with the identity level about 68–74% to the standard copy); (ii) genes encoding cation transporter and methyltransferase; (iii) genes encoding transposases of different IS elements.

To our knowledge, such complex operons of copper resistance have not yet been studied in detail.

3.3.4. Determinants of Arsenic Resistance. To overcome the toxic effect of arsenic compounds bacteria use different mechanisms based on their active extrusion, extracellular precipitation and chelation, and intracellular transformation [30]. Several genetic systems encoding arsenic resistance in Gram-negative bacteria were revealed and studied to date [30, 31]. We studied in detail the structure of plasmids of two permafrost *A. lwoffii* strains, ED45-23 and ED5-9a, resistant to arsenic compounds (Table 2). Indeed, in each of these strains we revealed a single plasmid containing genes involved in arsenic resistance. These were plasmid pALWED2.1 (190039 bp) and plasmid pALWED3.1 (138027 bp).

In both plasmids we detected almost identical *ars* operons containing six genes: *trxB*, *arsH*, *arsB*, *arsC1*, *arsR*, and *arsC2* (Table 3 and Figure 1). Based on the analysis of published data it can be suggested that the genes of *ars* operon encode the following polypeptides: organoarsenic oxidase, thioredoxin reductase, oxyanion translocation protein, arsenate reductase, arsenite-inducible repressor, and second arsenate reductase protein [30, 32, 33].

The structure of *ars* operons involved in arsenic resistance in *A. lwoffii* was most similar to those (*arsH*, *arsB*, *arsC*, and *arsR*) described in *P. putida* (NC_002947) and in *Herminiimonas arsenicoxydans* [30]. A significant difference was the presence of two additional genes: *trxB* at the beginning of *ars* operon and the *arsC* at the end of this operon. It should be mentioned that to our knowledge such variants of the *ars* operons have not been described yet.

3.3.5. Determinants of Resistance to Co/Zn/Cd. One of the first studied plasmid-encoded transport systems directed on elimination of toxic cations is the *czc* efflux system [34]. The *czc* resistance operon was originally found and studied in plasmid of *A. eutrophus* CH34. It was shown that *czc* operon includes four genes encoding four proteins, *CzcA*, *CzcB*, *CzcC*, and *CzcD*, from which *CzcA* is the central protein with efflux capacities [35].

We revealed functionally active *czc* operons in four permafrost *A. lwoffii* strains: ED23-35, ED9-5a, EK30A, and VS15 (Table 2). The *czc* operon harbored by pALWVS1.1 included the four genes *czcD*, *czcA*, *czcB*, and *czcC* (Table 3 and Figure 1). Closely related *czc* operon was found in the strain EK30A on the plasmid pALWEK1.1 (identity about 97%). These two operons were for convenience designated *czc1*. Two other *czc* operons differed from *czc1* as well as one from another.

The *czc2* operon revealed on the plasmid pALWED1.1 (strain ED23-35) was only distantly related to *czc1* (about 50% identity at aa sequence level). In addition, between genes *czcD* and *czsA* it contained about 4 kb insertion with six ORFs, including three ones encoding the resistance to nickel (*rcnR*, *nreB*, and gene of Ni permease). Accordingly, strain ED23-35, as opposed to other *A. lwoffii* strains, was resistant to nickel salt (Table 3 and Figure 1).

The third *czc* operon, *czc3*, located on plasmid pALWED3.1 (strain ED9-5a) differed significantly from *czs1* as well as from *czc2*. This operon was most closely related (97% identity) to the chromosomal *czc* operon of *Moraxella osloensis* CCUG 350 (CP014234), bacterial strain belonging to another genus of Moraxellaceae family. This observation certainly suggests the involvement of horizontal gene transfer in the evolution of heavy metal resistance in bacteria.

It should be noted that the *czc2* operon as well as *chrA-chrB* genes located on plasmid pALWED1.1 differed significantly from respective determinants found on the other plasmids. It can be assumed that these differences are due to different functional characteristics of these plasmids: plasmid pALWED1.1 is conjugative, while the two others are not.

3.4. Distribution of Resistance Genes in Chromosomes and Plasmids of *A. lwoffii*. We determined the distribution of

the resistance determinants found in plasmids of permafrost strains in plasmids of clinical *A. lwoffii* strains and in chromosomes of environmental and clinical strains of *A. lwoffii*.

We focused on the analysis of contigs obtained in our work during the whole-genome sequencing of ancient *A. lwoffii* strains and of whole-genome shotgun sequences of clinical *A. lwoffii* strains obtained from public databases. It should be noted that, unlike *A. baumannii* strains, no complete genome sequence of *A. lwoffii* was published and all whole-genome shotgun sequences of this species belong to clinical strains.

We have found in the clinical strains the whole set of resistance determinants which we revealed on the plasmids of the permafrost strains (see Table S1 in Supplementary Material available online at <http://dx.doi.org/10.1155/2016/3970831>). However, there were significant differences between the determinants found on chromosomes and plasmids. The determinants revealed on the plasmids of clinical strains were closely related to those of the permafrost strains plasmids. The chromosomal determinants of all kinds/types formed distinct lineages and were characterized by different features in comparison with plasmid determinants (Table 4).

In particular, we detected on presumably plasmid contigs of clinical *A. lwoffii* strains *mer* operons, *ars* operons, *czc* operons, *cop* operons, and *chrA-chrB* genes almost identical (99-100% identity) to those located in the permafrost plasmids (Table S1). It is important to note also that in the plasmids of the clinical strains the different resistance determinants were linked to each other (Table S1). For instance, we have found in similar 14 kb contigs of the strains NIPH512 and NBRC 109760 resistance determinants to mercury, arsenic, and chromium and in the strain NIPH 715 (33 kb contig) resistance determinants to mercury, arsenic, and copper (in Table S1 these contigs are shown in bold). Therefore the linked location of operons involved in the resistance to different heavy metals can be regarded as a characteristic feature of *A. lwoffii* plasmids.

Resistance determinants located on bacterial chromosomes differed from the plasmids by several features: (i) the vast majority of chromosomal operons contained incomplete sets of genes or defective operons (the only exception is the *czc* operon and genes *chrA* and *chrB* from the strain CIP64.10 (APQS00000000)) (Table 4); (ii) the proteins encoded by chromosome and plasmid genes shared less than 85% identical amino acid residues (from 44% to 83%) and differed by their size (Table 4). Note that all these features are common to the environmental as well as to clinical *A. lwoffii* strains.

To illustrate this situation it is best to consider the data on the structure of chromosomal *ars* operons and their distribution in the chromosomes of *A. lwoffii* (Table S2). It is seen that, in many cases, chromosomal *ars* operons are incomplete (missing one or two from six genes) (Table S2). In other cases all genes are present, but apparently the operon is functionally nonactive due to nonsense or frame-shift mutations or deletions in some of its genes. To verify the functional activity of chromosomal *ars* operons we compared the resistance to arsenic of ancient *A. lwoffii* strains containing plasmid-encoded *ars* operon (ED45-23 and ED9-5a) and strains containing only chromosomal *ars* operons (ED23-35,

TABLE 4: Heavy metal resistance determinants in chromosomes and plasmids in strains of *A. lwoffii*.

| Strain, AC | Localization of genes | Presence of resistance genes and size of corresponding protein (aa) | | | | | | | | | |
|---|-----------------------|---|------|------|------|------|-----------|------|------|------|------|
| | | ChrA | ChrB | CzcA | CzcB | CzcC | CzcD | CopA | CopB | CopC | CopD |
| NCTC 5866 = CIP64.10 = NIPH 512, APQS00000000 | Chromosome | 401 | — | — | 408 | 441 | 305 + 313 | 588 | 319 | — | — |
| | Plasmid* | 449 | 308 | 1054 | — | — | — | 580 | 329 | 126 | 308 |
| NIPH 478, APQU00000000 | Chromosome | 401 | — | — | — | — | 313 | 588 | 314 | — | — |
| | Plasmid* | 449 | 308 | 1052 | 411 | 455 | 299 | nd | 306 | 126 | 308 |
| NIPH 715, APOT00000000 | Chromosome | 401 | — | — | — | — | 313 | 560 | 315 | — | — |
| | Plasmid* | — | — | 1052 | 416 | 455 | 284 | 603 | 339 | 126 | 308 |
| ATCC 9957 = CIP 70.31, APQT00000000 | Chromosome | 401 | — | — | — | — | 313 | 588 | 314 | — | — |
| | Plasmid* | 449 | 308 | 1052 | 411 | 455 | 299 | nd | 306 | 126 | 308 |
| TG19636, AMJG00000000 | Chromosome | 401 | — | — | — | — | 259 | 588 | 339 | — | — |
| | Plasmid* | — | — | 1052 | 416 | 455 | 305 | 617 | 306 | 126 | 308 |
| EJ23-35 | Chromosome | 401 | 401 | — | — | — | 259 | 580 | 339 | — | — |
| | Plasmid | 449 | 308 | 1042 | 491 | 399 | 315 | — | — | — | — |
| VS15 | Chromosome | 401 | — | — | — | — | — | 585 | 315 | — | — |
| | Plasmid | — | — | 1052 | — | — | — | 637 | 306 | 126 | 308 |
| EK30A | Chromosome | 401 | — | — | — | — | 259 | 585 | 315 | — | — |
| | Plasmid | 449 | 308 | 1052 | 436 | 455 | 299 | 637 | 306 | 126 | 308 |

* Contig identified as a plasmid with a high probability; nd: not determined.

VS15, and EK30a). All three strains with only chromosomal *ars* genes were sensitive to arsenic, unlike strains containing plasmid operons (Table 2). Thus, ancient *A. lwoffii* strains contain active *ars* operons only in their plasmids. However, it can be suggested that chromosomal *ars* operons played an important role in the evolution of *A. lwoffii*, because of their presence in multiple copies on chromosomes in ancient as well as in present-day strains of this species.

It should be noted that other chromosomal operons (*cop*, *czc*, and *chrA-chrB*), as a rule, also did not contain a complete set of genes. The *mer* operon was an exception since we failed to detect its chromosomal copies in any of the studied strains of *A. lwoffii*.

3.5. Comparative Analysis of Genomes of *A. lwoffii* and *A. baumannii*. Our data indicate that functionally active operons of *A. lwoffii* encoding resistance to heavy metal salts are preferably located on plasmids and that many strains of *A. lwoffii* are characterized by the presence of numerous plasmids, including plasmids larger than 100 kb. To determine whether these properties are also inherent to clinical *Acinetobacter* strains, especially to strains of *A. baumannii*, we compared genomic sequences of *A. lwoffii* and *A. baumannii*, with particular attention to the size of their chromosomes and the number and structure of their plasmids.

For our comparative analysis, we selected a collection of well-studied strains of *A. baumannii* described in the work of Ou et al. [36].

Based on the phylogenetic data, three of the strains (AYE, AB307-0294, and AB0057) belonged to global clone I (GC I). Two of the three BJAB strains (BJAB07104 and BJAB0868), along with 4 previously reported Asia strains, including MDR-ZJ06 (China), MDR-TJ (China), ABTCD0715

(Taiwan), and ABI656-2 (Korean), were grouped together with ACICU, a strain of global clone II (GC II) group. The strain BJAB0715 probably has a different origin in comparison with other drug-resistant strains. Strains ATCC17978 and AB307-0294 were susceptible to antibiotics and strain SDF was isolated from a human body louse [32, 36].

In addition, we analyzed a number of *A. baumannii* strains whose genomes were completely sequenced and studied in recent years. In total, we analyzed the genomes of 31 strains of *A. baumannii*.

Unlike *A. baumannii*, *A. lwoffii* strains have only recently begun to be actively studied. To date, information on the whole-genome shotgun sequencing of only 7 clinical isolates of *A. lwoffii* was deposited to database, but the authors did not perform their analysis. In our work we complemented the existing data by sequencing of the genomes of five ancient *A. lwoffii* strains isolated from permafrost deposits.

Analysis of clinical *A. lwoffii* genomes demonstrated that average size of the genomes of the present-day clinical strains is about 3.4 Mbp (Table S3). According to our data, the genomes of permafrost *A. lwoffii* strains have similar sizes.

Comparative analysis of well-studied strains of *A. baumannii* demonstrated that they have genome size of about 4.0 Mbp and contain no more than 4 plasmids (Table S4). Despite the differences in the number and quality of the genomic sequences of the two species, the results clearly indicate that the dimensions of the *A. baumannii* chromosomes exceeded those of *A. lwoffii* by about 20%. Less reliable conclusion is possible about the differences in the number of plasmids present in the strains of *A. baumannii* and *A. lwoffii* due to the limited data available. It seems that the plasmids in the *A. lwoffii* strains are more numerous than in *A. baumannii*.

At present, many plasmids found in *A. baumannii* strains are identified and sequenced. We have analyzed the structure of 34 medium-size and large plasmids but detected only one plasmid (pD36-4) containing full *mer* operon and one with fragments of *ars* operon (pAB3). At the same time, we detected full resistance operons in chromosomes of many strains of this species. For instance, we revealed *mer* and *ars* operons in the chromosomes of strains AYE and AB0057 [3]; copper operon in the chromosomes of strains LAC-4, ATCC 17978, and AB0057 [31]; *czc* operons in the chromosomes of strains ATCC17978 (CP000521), ACICU (CP000863), AB5075-UW (CP008706), AB0057 (CP001182), and AYE (CU459141).

Thus, unlike *A. baumannii* plasmids, plasmids of *A. lwoffii* carry a big load of functionally active heavy metal resistance determinants; this corresponds to larger numbers and sizes of plasmids in *A. lwoffii* compared to *A. baumannii* (Table 2 and Table S4).

The nature of these differences, whether or not they are associated with lifestyles of two species in distinct ecological niches, is a subject for further studies.

4. Conclusions

We for the first time sequenced and analyzed several large plasmids (130–280 kb), which encode genes multiresistance to heavy metals and arsenic compounds, in environmental strains of *A. lwoffii*, a potential human pathogen. All five large plasmids of *A. lwoffii* contained genes of heavy metal resistance in different combinations. In particular, we revealed genes of resistance to mercury, arsenate, chromium, copper, nickel, and cobalt/zinc/cadmium. The *cop* loci and *ars* operon differed by their structure from those described earlier. *Mer* and *ars* operons were linked to each other. We also obtained preliminary data showing that large plasmids encoding for multiresistance to heavy metals are present in clinical isolates of *A. lwoffii*. Thus it seems that in *A. lwoffii* determinants of heavy metal resistance are often located on plasmids.

A completely different situation was revealed in studies of metal-resistant *A. baumannii* strains. We have analyzed completely sequenced genomes of 31 *A. baumannii* strains from public databases including their chromosomes as well as plasmids. As a rule, heavy metal resistance genes forming full operons were located on chromosomes and only sometimes were they found on plasmids. In most cases, chromosomal resistance genes were related to the plasmid genes found in *A. lwoffii* strains. Furthermore, the average chromosome size of *A. baumannii* exceeded that of *A. lwoffii* by about 20%, which might indicate the higher role of chromosomes of *A. baumannii* in their lifestyle in comparison with *A. lwoffii*.

The results of our work suggest the occurrence of two different strategies in two species of the same genus revealed by different structural and functional roles of plasmids and chromosomes. One can speculate that these two strategies may be due to the fact that *A. lwoffii* strains, unlike *A. baumannii*, mainly inhabit natural environments. Furthermore, our data support hypothesis that not only mercury but also other heavy metals and arsenic have great impact on the evolution of *Acinetobacter* genomes. The accuracy of these

conclusions should be confirmed by further studies of plasmids harbored by both environmental and clinical strains of *A. lwoffii*.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This research was partially supported by the Russian Foundation for Basic Research, Grants 14-04-01917 and 14-04-01317, and by the Russian Academy of Sciences Presidium Program “Molecular and Cellular Biology” (grants to A. Kulbachinsky and A. Mardanov). The authors are grateful to A. Kulbachinsky for helpful comments and suggestions and for critical reading of the manuscript.

References

- [1] M. Shintani, Z. K. Sanchez, and K. Kimbara, “Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy,” *Frontiers in Microbiology*, vol. 6, article 242, 2015.
- [2] J. M. Coombs and T. Barkay, “New findings on evolution of metal homeostasis genes: evidence from comparative genome analysis of bacteria and archaea,” *Applied and Environmental Microbiology*, vol. 71, no. 11, pp. 7083–7091, 2005.
- [3] P.-E. Fournier, D. Vallenet, V. Barbe et al., “Comparative genomics of multidrug resistance in *Acinetobacter baumannii*,” *PLoS Genetics*, vol. 2, no. 1, article e7, 2006.
- [4] M. S. Moghadam, A. Albersmeier, A. Winkler et al., “Isolation and genome sequencing of four Arctic marine *Psychrobacter* strains exhibiting multicopper oxidase activity,” *BMC Genomics*, vol. 17, no. 1, article 117, 2016.
- [5] M. Touchon, J. Cury, E. J. Yoon et al., “The genomic diversification of the whole acinetobacter genus: origins, mechanisms, and consequences,” *Genome Biology and Evolution*, vol. 6, no. 10, pp. 2866–2882, 2014.
- [6] D. Vallenet, P. Nordmann, V. Barbe et al., “Comparative analysis of acinetobacters: three genomes for three lifestyles,” *PLoS ONE*, vol. 3, no. 3, Article ID e1805, 2008.
- [7] L. Zhu, Z. Yan, Z. Zhang et al., “Complete genome analysis of three *Acinetobacter baumannii* clinical isolates in China for insight into the diversification of drug resistance elements,” *PLoS ONE*, vol. 8, no. 6, Article ID e66584, 2013.
- [8] M. Fondi, G. Bacci, M. Brilli et al., “Exploring the evolutionary dynamics of plasmids: the *Acinetobacter* pan-plasmidome,” *BMC Evolutionary Biology*, vol. 10, article 59, 2010.
- [9] G. Kholodii, S. Mindlin, Z. Gorlenko, M. Petrova, J. Hobman, and V. Nikiforov, “Translocation of transposition-deficient (TndPKLH2-like) transposons in the natural environment: mechanistic insights from the study of adjacent DNA sequences,” *Microbiology*, vol. 150, no. 4, pp. 979–992, 2004.
- [10] M. A. Petrova, S. Z. Mindlin, Zh. M. Gorlenko, E. S. Kaliaeva, V. S. Soina, and E. S. Bogdanova, “Mercury-resistant bacteria from permafrost sediments and prospects for their use in comparative studies of mercury resistance determinants,” *Genetika*, vol. 38, no. 11, pp. 1569–1574, 2002 (Russian).

- [11] J. Sambrook and D. W. Russel, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA, 3rd edition, 2001.
- [12] S. F. Altschul, T. L. Madden, A. A. Schäffer et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389–3402, 1997.
- [13] R. J. Roberts, T. Vincze, J. Posfai, and D. Macelis, "REBASE-A database for DNA restriction and modification: enzymes, genes and genomes," *Nucleic Acids Research*, vol. 38, no. 1, pp. D234–D236, 2009.
- [14] A. Marchler-Bauer, S. Lu, J. B. Anderson et al., "CDD: a Conserved Domain Database for the functional annotation of proteins," *Nucleic Acids Research*, vol. 39, supplement 1, pp. D225–D229, 2011.
- [15] R. D. Finn, J. Mistry, J. Tate et al., "The Pfam protein families database," *Nucleic Acids Research*, vol. 38, no. 1, pp. D211–D222, 2009.
- [16] K. Hirai, H. Aoyama, S. Suzue, T. Irikura, S. Iyobe, and S. Mitsuhashi, "Isolation and characterization of norfloxacin-resistant mutants of *Escherichia coli* K-12," *Antimicrobial Agents and Chemotherapy*, vol. 30, no. 2, pp. 248–253, 1986.
- [17] A. Kurakov, S. Mindlin, A. Beletsky et al., "The ancient small mobilizable plasmid pALWED1.8 harboring a new variant of the non-cassette streptomycin/spectinomycin resistance gene *aadA27*," *Plasmid*, vol. 84–85, pp. 36–43, 2016.
- [18] C. Große, G. Grass, A. Anton et al., "Transcriptional organization of the *czc* heavy-metal homeostasis determinant from *Alcaligenes eutrophus*," *Journal of Bacteriology*, vol. 181, no. 8, pp. 2385–2393, 1999.
- [19] T. Barkay, S. M. Miller, and A. O. Summers, "Bacterial mercury resistance from atoms to ecosystems," *FEMS Microbiology Reviews*, vol. 27, no. 2–3, pp. 355–384, 2003.
- [20] C. Cervantes, H. Ohtake, L. Chu, T. K. Misra, and S. Silver, "Cloning, nucleotide sequence, and expression of the chromate resistance determinant of *Pseudomonas aeruginosa* plasmid pUM505," *Journal of Bacteriology*, vol. 172, no. 1, pp. 287–291, 1990.
- [21] A. Nies, D. H. Nies, and S. Silver, "Nucleotide sequence and expression of a plasmid-encoded chromate resistance determinant from *Alcaligenes eutrophus*," *The Journal of Biological Chemistry*, vol. 265, no. 10, pp. 5648–5653, 1990.
- [22] A. H. Alvarez, R. Moreno-Sánchez, and C. Cervantes, "Chromate efflux by means of the ChrA chromate resistance protein from *Pseudomonas aeruginosa*," *Journal of Bacteriology*, vol. 181, no. 23, pp. 7398–7400, 1999.
- [23] R. Francisco, M. C. Alpoim, and P. V. Morais, "Diversity of chromium-resistant and -reducing bacteria in a chromium-contaminated activated sludge," *Journal of Applied Microbiology*, vol. 92, no. 5, pp. 837–843, 2002.
- [24] N. L. Brown, D. A. Rouch, and B. T. O. Lee, "Copper resistance determinants in bacteria," *Plasmid*, vol. 27, no. 1, pp. 41–51, 1992.
- [25] C. Cervantes and F. Gutierrez-Corona, "Copper resistance mechanisms in bacteria and fungi," *FEMS Microbiology Reviews*, vol. 14, no. 2, pp. 121–137, 1994.
- [26] H. Basim, G. V. Minsavage, R. E. Stall, J.-F. Wang, S. Shanker, and J. B. Jones, "Characterization of a unique chromosomal copper resistance gene cluster from *Xanthomonas campestris* pv. *vesicatoria*," *Applied and Environmental Microbiology*, vol. 71, no. 12, pp. 8284–8291, 2005.
- [27] T. von Rozycki and D. H. Nies, "*Cupriavidus metallidurans*: evolution of a metal-resistant bacterium," *Antonie Van Leeuwenhoek*, vol. 96, no. 2, pp. 115–139, 2009.
- [28] S. P. Ng, E. A. Palombo, and M. Bhavé, "The heavy metal tolerant soil bacterium *Achromobacter* sp. AO22 contains a unique copper homeostasis locus and two mer operons," *Journal of Microbiology and Biotechnology*, vol. 22, no. 6, pp. 742–753, 2012.
- [29] S. Silver and G. Ji, "Newer systems for bacterial resistances to toxic heavy metals," *Environmental Health Perspectives*, vol. 102, supplement 3, pp. 107–113, 1994.
- [30] J. Andres and P. N. Bertin, "The microbial genomics of arsenic," *FEMS Microbiology Reviews*, vol. 40, no. 2, pp. 299–322, 2016.
- [31] S. Silver and L. T. Phung, "Bacterial heavy metal resistance: new surprises," *Annual Review of Microbiology*, vol. 50, pp. 753–789, 1996.
- [32] J. Ye, H.-C. Yang, B. P. Rosen, and H. Bhattacharjee, "Crystal structure of the flavoprotein ArsH from *Sinorhizobium meliloti*," *FEBS Letters*, vol. 581, no. 21, pp. 3996–4000, 2007.
- [33] J. Chen, H. Bhattacharjee, and B. P. Rosen, "ArsH is an organo-arsenical oxidase that confers resistance to trivalent forms of the herbicide monosodium methylarsenate and the poultry growth promoter roxarsone," *Molecular Microbiology*, vol. 96, no. 5, pp. 1042–1052, 2015.
- [34] S. Silver and M. Walderhaug, "Gene regulation of plasmid- and chromosome-determined inorganic ion transport in bacteria," *Microbiological Reviews*, vol. 56, no. 1, pp. 195–228, 1992.
- [35] A. Nies, D. H. Nies, and S. Silver, "Cloning and expression of plasmid genes encoding resistances to chromate and cobalt in *Alcaligenes eutrophus*," *Journal of Bacteriology*, vol. 171, no. 9, pp. 5065–5070, 1989.
- [36] H.-Y. Ou, S. N. Kuang, X. He et al., "Complete genome sequence of hypervirulent and outbreak-associated *Acinetobacter baumannii* strain LAC-4: epidemiology, resistance genetic determinants and potential virulence factors," *Scientific Reports*, vol. 5, article 8643, 2015.

Research Article

Molecular Cloning and Characterization of Full-Length cDNA of Calmodulin Gene from Pacific Oyster *Crassostrea gigas*

Xing-Xia Li,¹ Wen-Chao Yu,² Zhong-Qiang Cai,³ Cheng He,² Na Wei,² Xiao-Tong Wang,² and Xi-Qing Yue¹

¹College of Food Science, Shenyang Agricultural University, Shenyang 110866, China

²School of Agriculture, Ludong University, Yantai 264025, China

³Changdao Enhancement and Experiment Station, Chinese Academy of Fishery Sciences, Changdao 265800, China

Correspondence should be addressed to Xiao-Tong Wang; wangxiaotong999@163.com and Xi-Qing Yue; yxq-nydx@163.com

Received 6 January 2016; Revised 16 June 2016; Accepted 28 June 2016

Academic Editor: William H. Piel

Copyright © 2016 Xing-Xia Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The shell of the pearl oyster (*Pinctada fucata*) mainly comprises aragonite whereas that of the Pacific oyster (*Crassostrea gigas*) is mainly calcite, thereby suggesting the different mechanisms of shell formation between above two mollusks. Calmodulin (CaM) is an important gene for regulating the uptake, transport, and secretion of calcium during the process of shell formation in pearl oyster. It is interesting to characterize the CaM in oysters, which could facilitate the understanding of the different shell formation mechanisms among mollusks. We cloned the full-length cDNA of Pacific oyster CaM (cgCaM) and found that the cgCaM ORF encoded a peptide of 113 amino acids containing three EF-hand calcium-binding domains, its expression level was highest in the mantle, hinting that the cgCaM gene is probably involved in shell formation of Pacific oyster, and the common ancestor of Gastropoda and Bivalvia may possess at least three CaM genes. We also found that the numbers of some EF hand family members in highly calcified species were higher than those in lowly calcified species and the numbers of these motifs in oyster genome were the highest among the mollusk species with whole genome sequence, further hinting the correlation between CaM and biomineralization.

1. Introduction

Calmodulin (CaM) seems to play important roles in regulating the uptake, transport, and secretion of calcium in molluscan shell formation, such as yesso scallop (*Patinopecten yessoensis*) [1], freshwater pearl mussel (*Hyriopsis schlegelii*) [2], and pearl oyster (*Pinctada fucata*) [3]. The pearl oyster CaM protein or calmodulin-like protein (CaLP) was found to interact with different target proteins in the mantle and the gill [4] and may regulate the calcite growth in the shell prismatic layer by inducing the aragonite nucleation and modifying the calcite morphology in vitro crystallization experiments [5]. The pearl oyster shell mainly comprises aragonite [6, 7], whereas the Pacific oyster shell is mainly calcite [8]. The higher level of calcite in the oyster shell may suggest that CaM has a different function in the Pacific oyster.

To elucidate the role of CaM in shell formation in the Pacific oyster, we cloned the full-length cDNA of Pacific oyster CaM (cgCaM) and analyzed its characteristics.

2. Materials and Methods

2.1. Specimens, RNA Isolation, and Primers. Six organ types (hemolymph, mantle, digestive gland, gill, adductor muscle, and labial palps) were collected from slaughtered Pacific oysters, snap-frozen in liquid nitrogen, and stored in the Ultra-low Temperature Freezer (−80°C) until extracting RNA. The total RNA was extracted using Trizol reagent (Tiangen, China). All of the primers were designed referring to one of the predicted CaM sequences (CGI_10006482) deposited in OysterBase (<http://www.oysterdb.com/>), as described in Table 1.

TABLE 1: Primers used for CaM cloning and expression analysis.

| Primer name | Sequence (5' → 3') | Size (nt) | Application |
|---------------------------------|--|-----------|--|
| 3 sites Adaptor Primer | CTGATCTAGAGGTACCGGATCC | 22 | 3' RACE |
| Oligo dT-3 sites Adaptor Primer | CTGATCTAGAGGTACCGGATCC(T) _n | 22+n | 3' RACE |
| GSP-3' RACE | GATGCTGATGGAAATGGAAC | 20 | 3' RACE |
| GSP1 | GGCTTCTCTCATCTCCTCTT | 20 | 5' RACE |
| GSP2 | CTTCTCTCATCTCCTCTTCG | 19 | 5' RACE |
| GSP3 | TTCTCTCATCTCCTCTTCGG | 19 | 5' RACE |
| AUAP | GGCCACGCGTCGACTAGTAC | 20 | 5' RACE |
| AAP | GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG | 38 | 5' RACE |
| CaM-F | GATGCTGATGGAAATGGAAC | 20 | For CaM gene in real-time PCR |
| CaM-R | TTCACCTCACTTCTTCCTCTG | 20 | For CaM gene in real-time PCR |
| eLFI-F | ACCACCCTGGTGAGATCAAG | 20 | For internal reference gene in real-time PCR |
| eLFI-R | ACGACGATCGCATTTCTCTT | 20 | For internal reference gene in real-time PCR |

2.2. Cloning the Full-Length cDNA. Full-length cgCaM mRNA sequences were got by 5'-RACE and 3'-RACE. For the 5'-RACE, the RACE-ready complementary DNA (cDNA) was got using SuperScript II (Invitrogen) and the primer GSP1; after synthesizing the cDNA, the deoxycytidine triphosphate and terminal deoxynucleotide transferase were used to add a poly-C tail to the 3' end of the cDNA; the primer pair of GSP2 and AAP was used to carry out the first PCR, the product of which was reamplified by GSP3 and AUAP according to the same thermal profile, which includes 2 minutes of an initial denaturation at 94°C, then 35 cycles comprising 30 seconds of denaturation at 94°C, 30 seconds of annealing at 55°C, and 2 minutes of extension at 72°C, 7 min of final extension at 72°C, and holding at 4°C at last. For the 3'-RACE, Oligo dT-3 sites Adaptor Primer (Takara) and AMV Reverse Transcriptase XL (Takara) were used to get the RACE-ready cDNA. The temperature profile employed was as follows: 10 minutes of at 30°C, 30 minutes at 50°C, 5 minutes at 95°C, and 5 minutes at 5°C. All of the RACE-ready cDNA products were amplified using sense primers (GSP-3' RACE) and antisense primers (3 sites Adaptor Primer, Takara), where the temperature profile was as follows: 5 minutes at 95°C; 30 cycles including 30 seconds at 94°C, 30 seconds at 55°C, and 5 minutes at 72°C; and then 7 minutes at 72°C. The PCR products of 5'-RACE and 3'-RACE were purified, ligated into the pMD18-T vector, and sequenced, respectively.

2.3. Expression Analysis of the cgCaM Gene in Different Organs. The mRNA levels of the cgCaM gene were measured using the method of Quantitative Real-time PCR. The quantities of cgCaM mRNA in different organs were standardized against that of elongation factor 1 (eLFI) mRNA to compensate for variations in volume of input mRNA. All of the reactions were carried out in a 20 µL volume, which

contained 10 µL of SYBR Green Master Mix (Tiangen), 5 pmol of each primer, 7 µL of water, and 1 µL of cDNA obtained by reverse transcription from each tissue (hemolymph, mantle, digestive gland, gill, adductor muscle, and labial palps) RNA using Moloney murine leukemia virus reverse transcriptase (Tiangen, China). Each cDNA sample was amplified with the primers CaM-F and CaM-R in three duplicates, each cDNA sample with the primers eLFI-F and eLFI-R in three duplicates, and the negative control in three duplicates on the same 96-well microplates. The PCR program was used according to the recommendation in the manufacturer. The specificity of the PCR products was evaluated by melting-curve analysis. The data were analyzed using the $2^{-\Delta\Delta Ct}$ method.

2.4. Sequence Character Analysis. The online tool, Compute pI/Mw, was used to predict the isoelectric point and molecular weight of cgCaM protein (<http://ca.expasy.org/tools/pi-tool.html>), and another online tool, the PSORT II program, was used to predict the N-terminal signal and cell location (<http://psort.ims.u-tokyo.ac.jp/>). Functional domain (IPR) prediction was performed at <http://prosite.expasy.org/>.

2.5. Phylogenetic Tree Construction. The CaM sequences of other mollusks were downloaded from NCBI and the CaM sequence from *Hydroides elegans* was used as an outgroup to root the tree. All the CaM sequences with complete CDS were got by the experimental methods in previous studies and downloaded from NCBI (<http://www.ncbi.nlm.nih.gov/>). A multiple sequence alignment was obtained using CLUSTALW. Phylogenetic tree was constructed using the neighbor-joining algorithm of MEGA 6 [9]. The robustness of the support for the tree was estimated

TGCCCTTATTCAAAGAGGCATTGAGCCTTTTGGACAAGGATGGAGATGGAACCATCACAACTAAAG

AACTGGGTACAGTT**ATG**AGATCCCTAGGACAAAATCCTACAGAGGCAGAGCTTCAAGACATG

ATTAACGAAGTTGATGCTGATGGAATGGAACCATAGATTTCCCGAATTTCTTACAATGA

TGGCAAAGAAAAATGAAGGACACTGATTCCGAAGAGGAGATGAGAGAAGCCTTCGGGGTGT

TCGATAAAGATGGAACGGTTTTATTAGTTGCGCAGAACTTCGTCACGTGATGACAAGTTT

GGGCGAGAGATTATCAGAGGAAGAAGTGAGTGAAATGATCAGGAAGCCGATATTGATGG

AGACGGTACCGTGAATTACGAAGAATTTGTCAAAATGATGACCAACAAG**TAG**AATGCGAAG

ATGCAATGTCGTAAAGAGCAGTTCAGATGTTATCAAATTATTTACGCAAAAAATTTATTATTGCAT

TCTTTTTAAAGATGTGACATATGTCATATAATAATGCATGATAAGATTAATGTGTTCTTTAGACTTTTT

TAGCGTTTATGAACATTGTGATAATAAACACGTGTCTTGAACATGAAGACATTGCAATAGAAAAA

AAAAAAAAAAAA

FIGURE 1: Sequence structure of cgCaM. The sequence highlighted in the upstream region indicates the 5' UTR and that in the downstream region is the 3' UTR. "ATG" in bold indicates the initiation codon and "TAG" in bold is the termination codon.

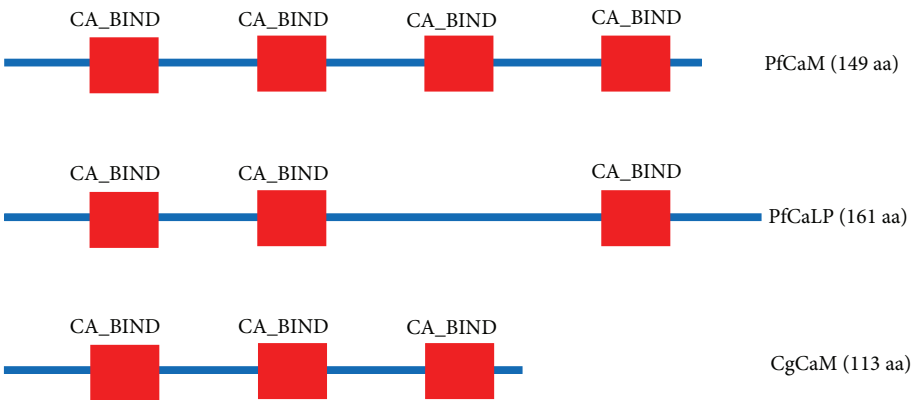


FIGURE 2: Functional domains of pfCaM, pfCaLP, and cgCaM. The InterProScan domain (IPR) was predicted at <http://prosite.expasy.org/>. The red block denotes the EF-hand calcium-binding domains (CA_BIND), which were located at amino acids 8–43, 45–80, and 81–113 in the Pacific oyster.

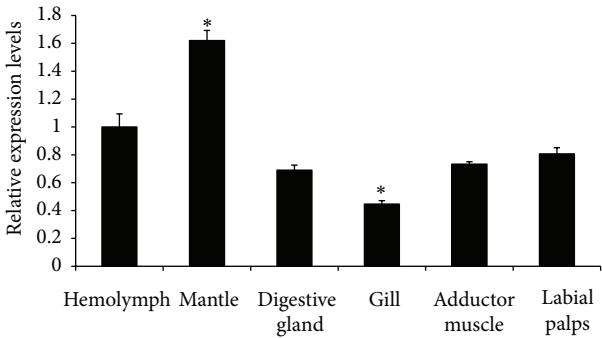


FIGURE 3: Relative expression levels of the cgCaM gene in different organs of Pacific oyster, that is, hemolymph, mantle, digestive gland, gill, adductor muscle, and labial palps. Asterisk (*) denoted that the cgCaM gene was expressed significantly highly ($p < 0.05$) in the mantle and significantly lowly in the gill ($p < 0.05$).

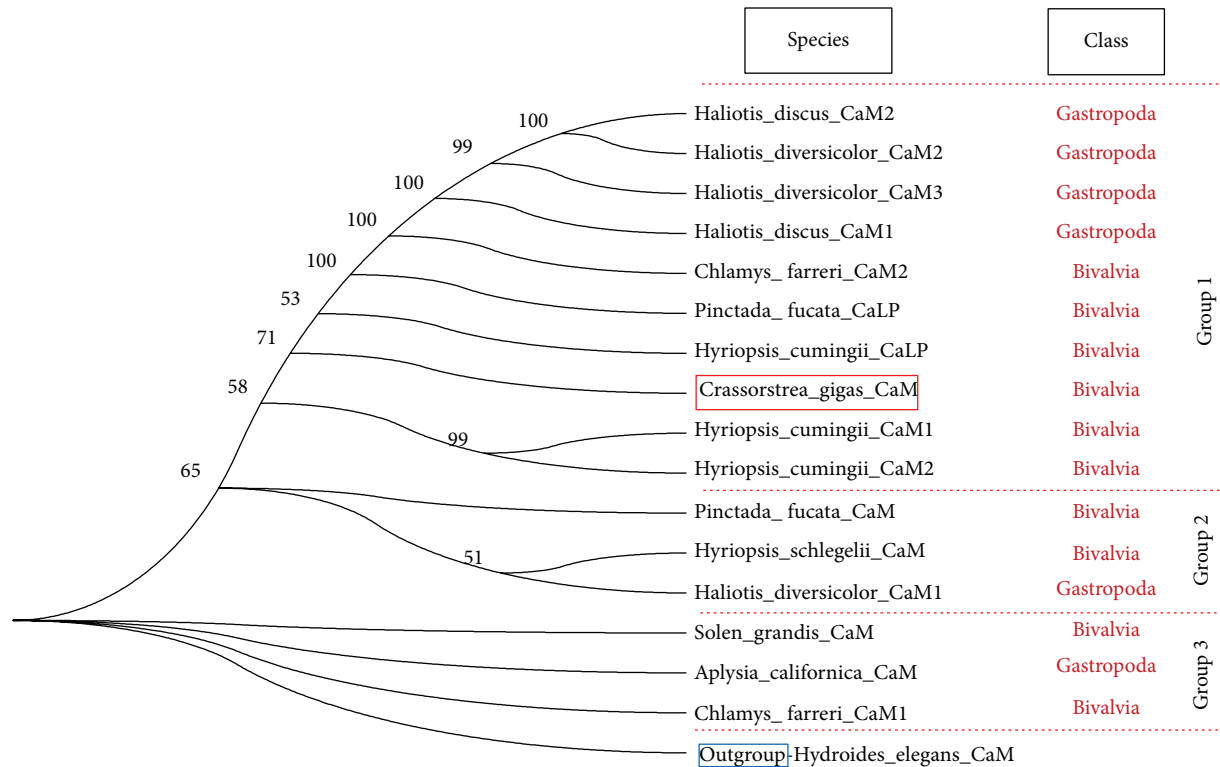


FIGURE 4: Phylogenetic tree of peptides orthologous to molluscan CaM. The peptides were divided into 3 groups including group 1, group 2, and group 3. CaM from *Hydroides elegans* was used as an outgroup.

based on 500 bootstrap replicates. In all of these analyses, the pairwise deletion option was used for handling alignment gaps as missing characters.

2.6. Calculation of the Numbers of EF Hand Clan Members in the Genomes of Multiple Species. We downloaded the Alignment file (Stockholm format) of the EF hand clan members from <http://pfam.xfam.org/clan/CL0220>, converted them into HMM file using hmmbuild tool in HMMER software and searched these motifs in the genomes of *Caenorhabditis elegans*, *Schistosoma ansoni*, *Schistosoma haematobium*, *Capitella teleta*, *Helobdella robusta*, *Culex quinquefasciatus*, *Drosophila melanogaster*, *Octopus bimaculoides*, *Daphnia magna*, *Daphnia pulex*, *Ciona intestinalis*, *Branchiostoma floridae*, *Callorhynchus milii*, *Saccoglossus kowalevskii*, *Aplysia californica*, *Biomphalaria glabrata*, *Pinctada fucata*, *Lingula anatina*, *Acropora digitifera*, *Strongylocentrotus purpuratus*, *Lottia gigantea*, *Homo sapiens*, *Gallus gallus*, *Crassostrea gigas*, *Chrysemys picta*, *Pelodiscus sinensis*, *Lepisosteus oculatus*, and *Salmo salar* using hmmsearch tool and calculated the motif numbers in these genomes. According to the degree of calcification, *Caenorhabditis elegans*, *Schistosoma ansoni*, *Schistosoma haematobium*, *Capitella teleta*, *Helobdella robusta*, *Culex quinquefasciatus*, *Drosophila melanogaster*, *Octopus bimaculoides*, *Daphnia magna* and *Daphnia pulex* were placed into the lowly calcified group; *Ciona intestinalis*, *Branchiostoma floridae*, *Callorhynchus milii*, *Saccoglossus kowalevskii*, *Aplysia*

californica, *Biomphalaria glabrata*, *Pinctada fucata*, *Lingula anatina*, *Acropora digitifera*, *Strongylocentrotus purpuratus*, and *Lottia gigantea* were placed into moderately calcified group; *Homo sapiens*, *Gallus gallus*, *Crassostrea gigas*, *Chrysemys picta*, *Pelodiscus sinensis*, *Lepisosteus oculatus*, and *Salmo salar* were placed into highly calcified group.

3. Results

3.1. Molecular Characteristics of cgCaM. The full-length cgCaM cDNA, which comprised 648 nucleotides, was obtained using the RACE technique. As shown in Figure 1, the ORF comprised 342 nucleotides and it encoded a peptide of 113 amino acids, the 5' UTR 80, and the 3' UTR 226 (GenBank accession number KM115543).

The cgCaM peptide was predicted to have an isoelectric point of 4.24 and a molecular weight of 12.86 kDa using the Compute pI/Mw tool. According to the PSORT II program, the cgCaM peptide was predicted to have no N-terminal signal peptide and it was expected to be located in the cytoplasm (probability = 60.9%), which is similar to the CaM and CaLP proteins in *Pinctada fucata* (pfCaM and pfCaLP).

After functional domain prediction, it was found that the cgCaM protein contained three EF-hand calcium-binding domain (CA_BIND) domains at amino acids 8–43, 45–80, and 81–113, and pfCaM also had three CA_BINDs (Figure 2). We did not identify an extra hydrophilic tail specific for CaLP or any signal peptide in cgCaM [10, 11].

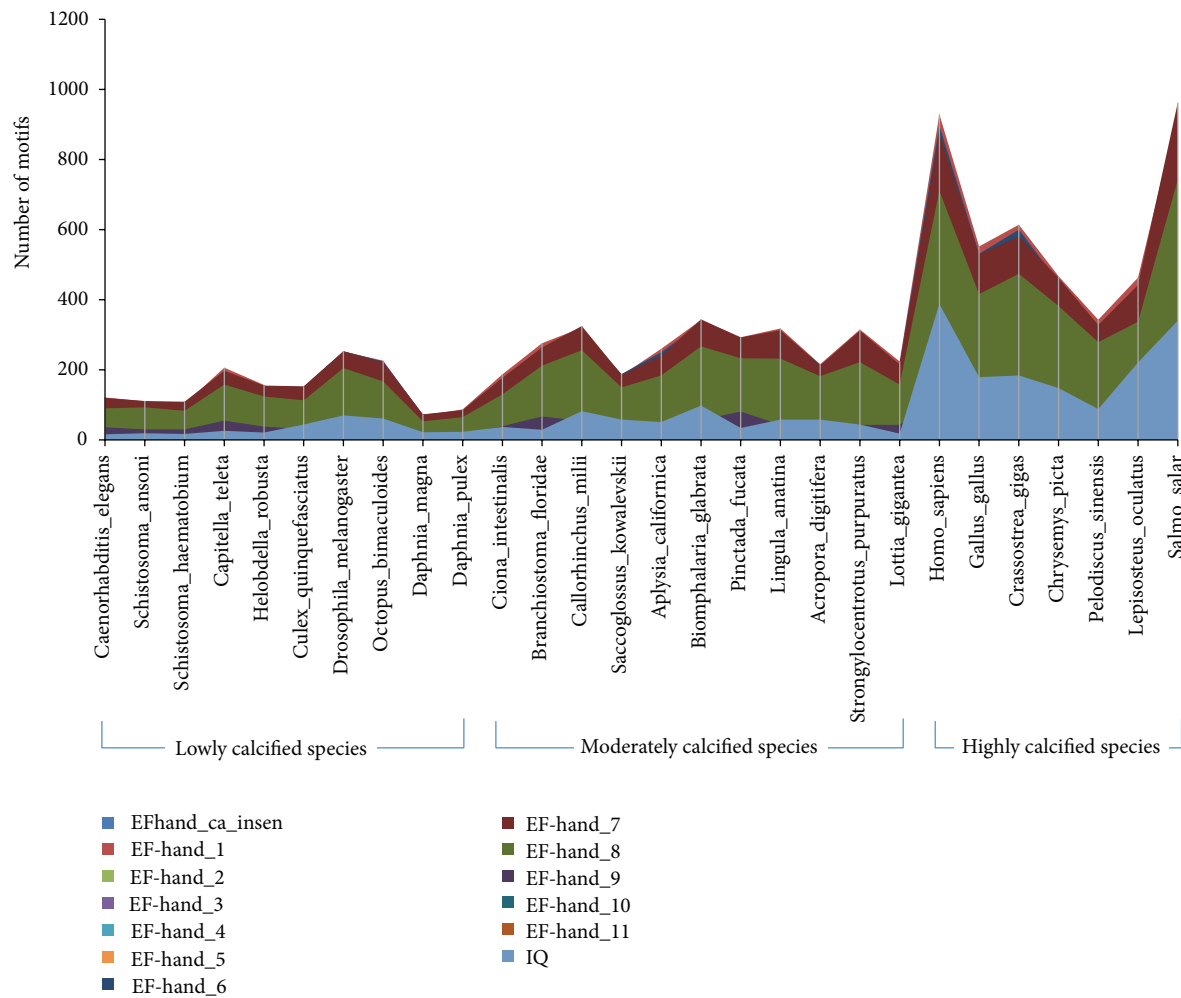


FIGURE 5: The distribution of EF hand motifs in lowly calcified, moderately calcified, and highly calcified species. EFhand.Ca_insen, EF-hand_1, EF-hand_2, EF-hand_3, EF-hand_4, EF-hand_5, EF-hand_6, EF-hand_7, EF-hand_8, EF-hand_9, EF-hand_10, EF-hand_11, EF-hand_like, and IQ belong to EF hand clan.

3.2. Expression Pattern of the cgCaM Gene in Different Organs. The mRNA expression levels of the cgCaM gene were examined by real-time PCR and compared indifferent organs (hemolymph, mantle, digestive gland, gill, adductor muscle, and labial palps). As shown in Figure 3, cgCaM was expressed in all six of the tissue types that we investigated, where the cgCaM mRNA expression level was highest in the mantle and lowest in the gill ($p < 0.05$).

3.3. Phylogenetic Tree of Peptides Orthologous to CaM. As shown in Figure 4, group 3 appeared earlier than group 2, and group 2 appeared earlier than group 1. Based on the phylogenetic tree of CaM and CaLP, we found that several Gastropoda and Bivalvia CaMs (group 3) were more similar with the outgroup CaM, whereas other molluscan CaMs (groups 1 and 2) were obviously different from the outgroup, while cgCaM appeared in group 1.

3.4. The Numbers of EF Hand Clan Members in the Genomes of Multiple Species. It was found that the numbers of some

EF hand family members, including EFhand.Ca_insen, EF-hand_1, EF-hand_2, EF-hand_3, EF-hand_4, EF-hand_5, EF-hand_6, EF-hand_7, EF-hand_8, EF-hand_9, EF-hand_10, EF-hand_11, EF-hand_like, and IQ, in highly calcified species were higher than those in lowly and moderately calcified species (Figure 5). Interestingly, the numbers of above motifs in oyster genome were the highest among the mollusk species with whole genome sequence (Figure 6).

4. Discussion

The shell is one of the most important features of mollusks because it is related to their health and growth. Thus, elucidating the mechanism of shell formation will be beneficial for mollusk culture. PfCaM and pfCaLP are believed to play important roles in shell formation of the pearl [10, 11] oyster and it was concluded that cgCaM may play a different role in shell formation by the Pacific oyster because the composition of the Pacific oyster shell is different from that of the pearl oyster. Determining the sequence characters,

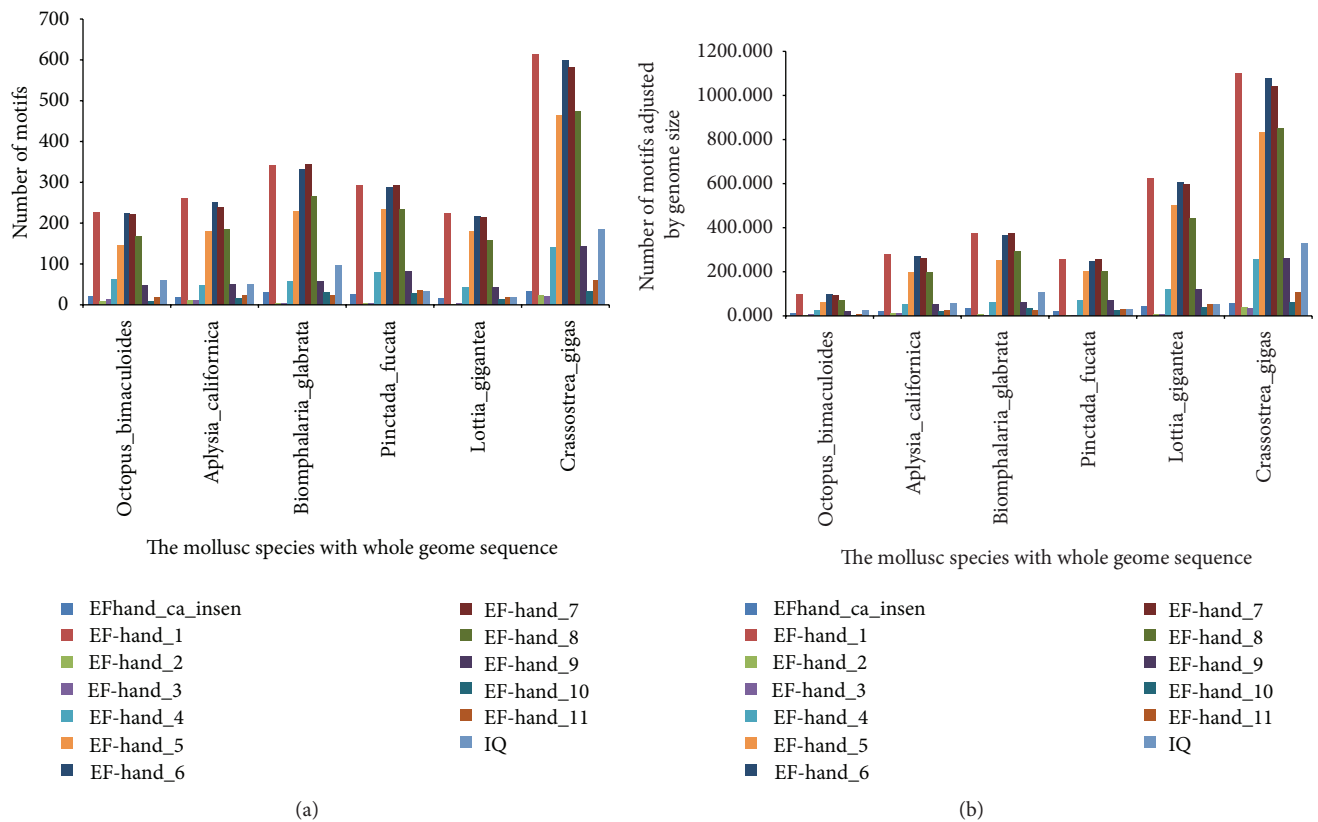


FIGURE 6: The numbers of some EF hand motifs in whole genomes of some mollusk species. (a) The motif numbers without adjustment; (b) the motif numbers adjusted using genome size.

expression pattern, and phylogenetic location of cgCaM will facilitate further research into shell formation by the Pacific oyster.

The mantle is an important organ for shell formation [12, 13], although some other organs may also be involved in the process [14, 15]. We found that the cgCaM gene expression level was highest in the mantle, thereby suggesting that its function is probably related to shell formation. However, CaM is present not only in mollusks but also in other animals without shell, such as *Hydroides elegans* [16] (Figure 4), which indicates that CaM has fundamental roles in calcium metabolism but not limited to shell formation and it may be involved in various pathways. Interestingly, we found that Gastropoda and Bivalvia CaMs were present in each branch (group) of the molluscan CaM tree (Figure 4), which suggests that the common ancestor of Gastropoda and Bivalvia probably had at least three CaM genes. In addition, the CaLP genes were only present in group 1 and group 1 was far away from the outgroup, which indicates that the CaLP gene emerged later than the CaM genes.

It was found that highly calcified species had more EF hand clan members in their genomes (Figure 5), hinting the probable correlation between these motifs and biomineralization. As we all know, oyster had the large and thick calcified shells [14, 15]; we also found more EF hand clan members in its genome than in other mollusk genomes (Figure 6), so it could be deduced that more EF hand clan members may

promote the formation of large and thick calcified shells. CaM mainly contained three motifs, EF-hand_1, EF-hand_2 and EF-hand_7, and may be involved in shell formation of other mollusks [1–3]; cgCaM also contained these three motifs and expressed highly in mantle (Figure 3), all of which hinted that cgCaM may be involved in oyster shell formation.

Competing Interests

All authors have not declared conflict of interests.

Authors' Contributions

Xing-Xia Li and Wen-Chao Yu contributed equally to this work.

Acknowledgments

This research was supported by National Natural Science Foundation of China (no. 31302181), Shandong Provincial Natural Science Foundation, China (no. ZR2013CM026), Key R & D program of Shandong Province (2015GSF115013), Modern Agricultural Industry Technology System of Shandong province, China (SDAIT-14-03), and Enterprise project (no. 2013HX007).

References

- [1] X. Sun, A. Yang, B. Wu, L. Zhou, and Z. Liu, "Characterization of the mantle transcriptome of yesso scallop (*Patinopecten yessoensis*): identification of genes potentially involved in biomineralization and pigmentation," *PLoS ONE*, vol. 10, no. 4, Article ID e0122967, 2015.
- [2] L.-G. Zeng, J.-H. Wang, Y.-J. Li, J.-Q. Sheng, Q. Gu, and Y.-J. Hong, "Molecular characteristics and expression of calmodulin cDNA from the freshwater pearl mussel, *Hyriopsis schlegelii*," *Genetics and Molecular Research*, vol. 11, no. 1, pp. 42–52, 2012.
- [3] S. Li, L. Xie, C. Zhang, Y. Zhang, M. Gu, and R. Zhang, "Cloning and expression of a pivotal calcium metabolism regulator: calmodulin involved in shell formation from pearl oyster (*Pinctada fucata*)," *Comparative Biochemistry and Physiology—B Biochemistry and Molecular Biology*, vol. 138, no. 3, pp. 235–243, 2004.
- [4] S. Li, L. Xie, Z. Ma, and R. Zhang, "cDNA cloning and characterization of a novel calmodulin-like protein from pearl oyster *Pinctada fucata*," *The FEBS Journal*, vol. 272, no. 19, pp. 4899–4910, 2005.
- [5] Z. Yan, Z. Fang, Z. Ma et al., "Biomineralization: functions of calmodulin-like protein in the shell formation of pearl oyster," *Biochimica et Biophysica Acta (BBA)—General Subjects*, vol. 1770, no. 9, pp. 1338–1344, 2007.
- [6] Z. Ma, J. Huang, J. Sun et al., "A novel extrapallial fluid protein controls the morphology of nacre lamellae in the pearl oyster, *Pinctada fucata*," *The Journal of Biological Chemistry*, vol. 282, no. 32, pp. 23253–23263, 2007.
- [7] K. Saruwatari, T. Matsui, H. Mukai, H. Nagasawa, and T. Kogure, "Nucleation and growth of aragonite crystals at the growth front of nacres in pearl oyster, *Pinctada fucata*," *Biomaterials*, vol. 30, no. 16, pp. 3028–3034, 2009.
- [8] S.-W. Lee, Y.-N. Jang, and J.-C. Kim, "Characteristics of the aragonitic layer in adult oyster shells, *Crassostrea gigas*: structural study of myostracum including the adductor muscle scar," *Evidence-Based Complementary and Alternative Medicine*, vol. 2011, Article ID 742963, 10 pages, 2011.
- [9] K. Tamura, G. Stecher, D. Peterson, A. Filipski, and S. Kumar, "MEGA6: molecular evolutionary genetics analysis version 6.0," *Molecular Biology and Evolution*, vol. 30, no. 12, pp. 2725–2729, 2013.
- [10] Z. Fang, Q. Wang, W. Cao et al., "Investigation of phosphorylation site responsible for CaLP (*P. fucata*) nucleo-cytoplasmic shuttling triggered by overexpression of p21Cip1," *Marine Biotechnology*, vol. 11, no. 2, pp. 270–279, 2009.
- [11] S. Li, L. Xie, Q. Meng, and R. Zhang, "Significance of the extra C-terminal tail of CaLP, a novel calmodulin-like protein involved in oyster calcium metabolism," *Comparative Biochemistry and Physiology B: Biochemistry and Molecular Biology*, vol. 144, no. 4, pp. 463–471, 2006.
- [12] F. Marin, G. Luquet, B. Marie, and D. Medakovic, "Molluscan shell proteins: primary structure, origin, and evolution," *Current Topics in Developmental Biology*, vol. 80, pp. 209–276, 2007.
- [13] T. Furuhashi, I. Miksik, M. Smrz et al., "Comparison of aragonitic molluscan shell proteins," *Comparative Biochemistry and Physiology B: Biochemistry and Molecular Biology*, vol. 155, no. 2, pp. 195–200, 2010.
- [14] G. Zhang, X. Fang, X. Guo et al., "The oyster genome reveals stress adaptation and complexity of shell formation," *Nature*, vol. 490, pp. 49–54, 2012.
- [15] X. Wang, L. Li, Y. Zhu et al., "Oyster shell proteins originate from multiple organs and their probable transport pathway to the shell formation front," *PLoS ONE*, vol. 8, no. 6, Article ID e66522, 2013.
- [16] Z.-F. Chen, H. Wang, and P.-Y. Qian, "Characterization and expression of calmodulin gene during larval settlement and metamorphosis of the polychaete *Hydroides elegans*," *Comparative Biochemistry and Physiology B: Biochemistry and Molecular Biology*, vol. 162, no. 4, pp. 113–119, 2012.

Research Article

Phylogeographic Structure of a Tethyan Relict *Capparis spinosa* (Capparaceae) Traces Pleistocene Geologic and Climatic Changes in the Western Himalayas, Tianshan Mountains, and Adjacent Desert Regions

Qian Wang,^{1,2} Ming-Li Zhang,^{1,3} and Lin-Ke Yin¹

¹Key Laboratory of Biogeography and Bioresource in Arid Land, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, Xinjiang 830011, China

²University of Chinese Academy of Sciences, Beijing 100049, China

³Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

Correspondence should be addressed to Ming-Li Zhang; zhangml@ibcas.ac.cn and Lin-Ke Yin; yinlk@ms.xjb.ac.cn

Received 27 February 2016; Accepted 27 April 2016

Academic Editor: William H. Piel

Copyright © 2016 Qian Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Complex geological movements more or less affected or changed floristic structures, while the alternation of glacials and interglacials is presumed to have further shaped the present discontinuous genetic pattern of temperate plants. Here we consider *Capparis spinosa*, a xeromorphic Tethyan relict, to discuss its divergence pattern and explore how it responded in a stepwise fashion to Pleistocene geologic and climatic changes. 267 individuals from 31 populations were sampled and 24 haplotypes were identified, based on three cpDNA fragments (*trnL-trnF*, *rps12-rpl20*, and *ndhF*). SAMOVA clustered the 31 populations into 5 major clades. AMOVA suggests that gene flow between them might be restricted by vicariance. Molecular clock dating indicates that intraspecific divergence began in early Pleistocene, consistent with a time of intense uplift of the Himalaya and Tianshan Mountains, and intensified in mid-Pleistocene. Species distribution modeling suggests range reduction in the high mountains during the Last Glacial Maximum (LGM) as a result of cold climates when glacier advanced, while gorges at midelevations in Tianshan appear to have served as refugia. Populations of low-altitude desert regions, on the other hand, probably experienced only marginal impacts from glaciation, according to the high levels of genetic diversity.

1. Introduction

An essential purpose of biogeography is to comprehend the mechanisms resulting in the current distribution patterns of organisms during their evolutionary histories [1, 2]. Events since the late Paleogene, especially the Quaternary geologic and climatic changes, likely had a far-reaching influence on the spatial geographic structure and genetic diversification of temperate species [3, 4]. After the Oligocene, the inland Tethys or Paratethys Sea retreated, accompanied by an increasingly droughty climate, resulting in elaboration of the xerophytic Tethyan Flora [5, 6]. From the Neogene to Quaternary, with the increasing aridification and uplift of the Himalayas, Tibet Plateau, and Tianshan Mountains [7, 8],

the originally continuous arid Tethyan Flora gradually evolved into the North Temperate, Mediterranean, West Asian-Central Asian, Central Asian, and Himalayan elements of today [9, 10]. Complex geological movements more or less affected or changed floristic structures, while rapid climatic shifts likewise impacted plant life worldwide during this time. Significantly, the alternation of glacials and interglacials during the Pleistocene is presumed to have played a vital role in further shaping the present discontinuous genetic pattern for a majority of temperate plant species in the Northern Hemisphere [6, 11, 12].

Phylogeography, a discipline which began more than 20 years ago, has mostly considered the response of geographic distribution and genetic structure of current species to

historical events and the cycles of glaciation [13–16]. Chloroplast DNA (cpDNA), suitable for reconstructing phylogenetic patterns of plant species, is effective in exploring and estimating phylogeographical history of temperate plants caused by climatic change [17–20]. Recent scholars have concentrated on the refuge during the Last Glacial Maximum (LGM) and range expansion during inter- or postglacial periods [21–24]. With the most focus on mountain plants in Europe, North America, and East Asia [25–27], phylogeographic patterns and demographical history of eremophytes in arid Northwestern China are also receiving increasing attention [28–30].

We use *Capparis spinosa*, a typical xerophyte, to verify the hypothesis of Pleistocene historical influence through phylogeographic analysis and discuss how its spatial genetic structure responded to paleogeologic and paleoclimatic changes. The perennial creeping subshrub is a type member of a relatively large genus of the Capparaceae, with a natural distribution in the Mediterranean, West, and Central Asia [31, 32]. In arid Western China, there is only one single genus (*Capparis* Tourn. ex L.) with one single species (*C. spinosa* L.) belonging exclusively to this family [33]. *C. spinosa* ranges over the Western Himalayas, Eastern Pamir Plateau, Tianshan Mountains, Turpan-Hami Basin, and the Hexi Corridor. Previous studies on *Capparis* have paid more attention to autecology, phenology, quantitative morphology, and reproductive ecology [34–37]. Molecular studies related to the genetic variation of the species are little explored [38, 39] and have not considered genetic divergence from a phylogeographic point of view.

Here we employed three maternally inherited cpDNA spacers, *trnL-trnF*, *rps12-rpl20*, and *ndhF*, to trace historical divergence patterns of *C. spinosa*. We design to address the following concrete issues. (1) According to phylogenetic relationships yielded by cpDNA sequence variation, what is the genetic pattern of the species? (2) When and how have allopatric divergence and genetic differentiation of this xerophyte been influenced by glacial-interglacial cycles associated with geologic and climatic changes?

2. Materials and Methods

2.1. Population Sampling. A total of 267 individuals of *C. spinosa* from 31 natural populations were investigated in the present study (Table 1), throughout almost the entire geographic distribution of the species in Eastern Central Asia, including Xinjiang, western Gansu, and western Tibet. For each population, leaf samples were collected from 5 to 12 mature and large perennial individuals separated by at least 100 m, in order to avoid inbreeding and clones. The latitude, longitude, and altitude at each collection center were measured using a global positioning system (GPS) receiver. About 10 g of fresh leaf materials was gathered per individual, dried in silica gel in the field, and stored at -20°C in the laboratory. *Capparis bodinieri* Lévl. was chosen as an outgroup for the phylogenetic study, based on previous taxonomic studies on *Capparis* [31]. Voucher specimens for all individuals were deposited in the Herbarium of the Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences (XJBI).

2.2. DNA Extraction, Amplification, and Sequencing. Silica-gel-dried leaf tissues were ground into powder in liquid nitrogen with a Fastprep-24 device (Sample Preparation System, MP Biomedicals, USA). Total genomic DNA was extracted from about 50 mg of powdered tissues using a modified 2x CTAB method [40]. Through preliminary screening from eighteen pairs of primers (see Table S1 in Supplementary Material available online at <http://dx.doi.org/10.1155/2016/5792708>), three cpDNA regions, *trnL-trnF*, *rps12-rpl20*, and *ndhF* (329F, 927R), were found because they have more sequence variation in individuals among populations. Polymerase chain reaction (PCR) amplification and DNA sequencing were performed using primer pairs for these three regions. PCRs were carried out in 30 μL reaction volumes including 1.5 μL of 10x buffer, 2 μL of 25 mmol/L MgCl_2 , 2.2 μL of 50 ng/ μL each primer (Sangon, Shanghai, China), 3 μL of 2.5 mmol/L dNTP, 0.3 μL of 5 U/ μL Taq DNA polymerase, and 1.0 μL of 5 ng/ μL genomic DNA. PCR amplification was run on a Gene-Amp PCR system 9700 DNA Thermal Cycler (Applied Biosystems, Foster City, CA, USA) with parameters set as follows: an initial start at 94°C with 4 min; followed by 30 cycles of 94°C with 30 s, then 59°C , 52°C , and 52°C , respectively, for *trnL-trnF*, *rps12-rpl20*, and *ndhF* with 30 s and 72°C with 90 s; plus a subsequent additional extension at 72°C with 10 min. Sizes of PCR products were visualized on 1% TAE-agarose gel electrophoresis. All PCR products were purified with PCR product purification kit (Shanghai SBS, Biotech Ltd., China) and then sequenced in both directions for *trnL-trnF*, *rps12-rpl20*, and *ndhF* using an ABI Prism 3730 xl automated sequencer in Sangon Biological Engineering Technology & Service Co., Ltd., Shanghai, China.

2.3. Data Analysis. The chloroplast DNA sequences were edited by SeqMan (Lasergene, DNASTAR Inc., Madison, Wisconsin, USA). All sequences were aligned with Clustal X version 1.81 [41], then refined, and adjusted. All resulting cpDNA sequences identified by nucleotide variations, signifying different haplotypes, were deposited in GenBank. The corresponding accession numbers are KU940216–KU940218 for *trnL-trnF*, KU866560–KU866562 and KU940220 for *rps12-rpl20*, and KU866548–KU866558 and KU940221–KU940223 for *ndhF* (329F, 927R); and accessions numbers of *C. bodinieri* as the outgroup are KU940219 for *trnL-trnF*, KU866563 for *rps12-rpl20*, and KU866559 for *ndhF* (329F, 927R). Different cpDNA haplotypes were identified by DnaSP 5.0 [42]. Geographic distribution of detected haplotypes was mapped using ArcMap 9.3 (ESRI, Redlands, CA, USA). Phylogenetic relationships among cpDNA haplotypes were associated via a median-joining network (MJN) method [43], which was constructed using Network 4.6.1.0 (available at http://www.fluxus-engineering.com/sharenet_rn.htm).

To maximize the differences of haplotype composition between geographical locations, spatial analysis of molecular variance algorithm was conducted using SAMOVA 1.0 procedure. This method divided the sampled populations into inferred partition (K groups) based on genetic differentiation and geographical homogeneity [44]. We initially set the K values ranged from 2 to 12 by repeated simulated annealing

TABLE 1: Details of geographical locations, sample sizes, cpDNA haplotypes, and diversity indices for 31 populations of *C. spinosa*.

| Code | Population locality | Latitude/longitude | Altitude (m) | No. | Haplotype | H_d (\pm SD) | π (\pm SD) |
|---|---------------------|--------------------|--------------|-----|-----------------------|---------------------|---------------------|
| Overall | | | | 267 | | 0.8925 \pm 0.0075 | 0.0037 \pm 0.0019 |
| (I) The Western Himalayas group | | | | | | | |
| (1) ZD | Zanda, Tibet | 31°30'N/79°38'E | 3,598 | 9 | H22, H23, H24 | 0.4167 \pm 0.1907 | 0.0021 \pm 0.0013 |
| (2) QS | Qusum, Tibet | 31°29'N/79°48'E | 3,645 | 5 | H22 | 0 | 0 |
| (II) The Eastern Pamir group | | | | | | | |
| (3) YGS | Yengisar, Xinjiang | 38°43'N/76°14'E | 2,010 | 9 | H6 | 0 | 0 |
| (4) TXG | Taxkorgan, Xinjiang | 37°47'N/75°13'E | 3,694 | 5 | H6 | 0 | 0 |
| (5) AKT | Akto, Xinjiang | 38°59'N/75°32'E | 2,441 | 9 | H6 | 0 | 0 |
| (6) WQ | Wuqia, Xinjiang | 39°44'N/75°40'E | 1,945 | 10 | H6, H21 | 0.4667 \pm 0.1318 | 0.0006 \pm 0.0005 |
| (7) ATX | Artux, Xinjiang | 39°52'N/76°47'E | 1,303 | 10 | H6, H7 | 0.4667 \pm 0.1318 | 0.0006 \pm 0.0005 |
| (8) JS | Jiashi, Xinjiang | 39°56'N/78°01'E | 1,162 | 9 | H6 | 0 | 0 |
| (III) The Ili Valley group | | | | | | | |
| (9) GL | Gongliu, Xinjiang | 43°37'N/81°49'E | 710 | 9 | H7 | 0 | 0 |
| (10) YN | Yining, Xinjiang | 43°37'N/82°08'E | 750 | 9 | H7 | 0 | 0 |
| (11) BL | Bole, Xinjiang | 44°42'N/82°05'E | 470 | 9 | H7 | 0 | 0 |
| (12) WS | Wusu, Xinjiang | 44°19'N/84°19'E | 685 | 9 | H7, H12 | 0.5000 \pm 0.1283 | 0.0013 \pm 0.0009 |
| (IV) The north side of the Tianshan Mountains group | | | | | | | |
| (13) KRMV | Karamay, Xinjiang | 44°56'N/84°46'E | 296 | 5 | H12 | 0 | 0 |
| (14) TL | Toli, Xinjiang | 45°55'N/83°36'E | 684 | 5 | H17 | 0 | 0 |
| (15) SHZ | Shihezi, Xinjiang | 44°16'N/85°57'E | 356 | 5 | H11 | 0 | 0 |
| (16) SW | Shawan, Xinjiang | 44°19'N/85°35'E | 584 | 9 | H5 | 0 | 0 |
| (17) MNS | Manas, Xinjiang | 44°11'N/86°08'E | 642 | 8 | H5, H10 | 0.4286 \pm 0.1687 | 0.0002 \pm 0.0002 |
| (18) URMQ | Urumqi, Xinjiang | 43°44'N/86°57'E | 1,060 | 8 | H10, H11 | 0.4286 \pm 0.1687 | 0.0010 \pm 0.0007 |
| (19) JMS | Jimsar, Xinjiang | 43°55'N/89°07'E | 945 | 8 | H5 | 0 | 0 |
| (V) The south side of the Tianshan Mountains group | | | | | | | |
| (20) KP | Kalpin, Xinjiang | 40°35'N/79°36'E | 1,131 | 10 | H1, H5 | 0.3556 \pm 0.1591 | 0.0008 \pm 0.0006 |
| (21) WNS | Wensu, Xinjiang | 41°29'N/79°54'E | 1,296 | 9 | H1, H8, H20 | 0.7500 \pm 0.0786 | 0.0016 \pm 0.0010 |
| (22) BC | Baicheng, Xinjiang | 41°59'N/83°03'E | 1,369 | 8 | H1, H8 | 0.5714 \pm 0.0945 | 0.0003 \pm 0.0003 |
| (23) LT | Luntai, Xinjiang | 41°14'N/84°12'E | 921 | 10 | H8, H14 | 0.3556 \pm 0.1591 | 0.0011 \pm 0.0007 |
| (24) KRL | Korla, Xinjiang | 41°50'N/86°03'E | 1,072 | 8 | H1, H8, H13 | 0.6071 \pm 0.1640 | 0.0004 \pm 0.0004 |
| (25) HJ | Hejing, Xinjiang | 42°31'N/86°16'E | 1,403 | 10 | H5, H8 | 0.4667 \pm 0.1318 | 0.0012 \pm 0.0008 |
| (26) TKS | Toksun, Xinjiang | 42°49'N/88°38'E | 95 | 12 | H1, H5 | 0.5455 \pm 0.0615 | 0.0012 \pm 0.0008 |
| (27) TRP | Turpan, Xinjiang | 42°52'N/89°11'E | -84 | 12 | H1, H5, H11, H18, H19 | 0.8485 \pm 0.0586 | 0.0017 \pm 0.0010 |
| (28) SS | Shanshan, Xinjiang | 43°11'N/90°09'E | 1,043 | 9 | H10, H15, H16 | 0.7222 \pm 0.0967 | 0.0015 \pm 0.0009 |
| (29) HM | Hami, Xinjiang | 42°31'N/94°09'E | 773 | 9 | H9, H10, H11 | 0.7500 \pm 0.0786 | 0.0013 \pm 0.0009 |
| (30) DH | Dunhuang, Gansu | 40°25'N/94°44'E | 1,051 | 10 | H1, H2, H3 | 0.3778 \pm 0.1813 | 0.0005 \pm 0.0004 |
| (31) GZ | Guazhou, Gansu | 40°22'N/95°34'E | 1,133 | 10 | H1, H2, H4, H5 | 0.7333 \pm 0.1005 | 0.0011 \pm 0.0007 |

No.: number of sample individuals; H_d : haplotype diversity; π : nucleotide diversity.

process until largest F_{CT} index was obtained. However, when $K \geq 6$, at least one of the groups contained only one population, implying that information on the group structure would have been disappearing [45–48]. To maintain the grouping pattern and exclude configurations with a single population in any of the groups, finally valid range of K was set as 2–5.

Parameters of Nei's diversity, containing haplotype diversity (H_d) and nucleotide diversity (π), were computed in Arlequin 3.1 program [49, 50]. To evaluate genetic variation within populations, among populations within groups, and among groups, AMOVA (analysis of molecular variance) was performed in the program Arlequin version 3.1 [50], and significance tests were conducted using 1,000 permutations. The statistics of average gene diversity and population differentiation, that is, the average gene diversity within populations (H_S), total gene diversity across all populations (H_T), the number of substitution types (N_{ST}), and genetic differentiation over populations (G_{ST}), were calculated using Permut version 1.0 [51] (available at <http://www.pierroton.inra.fr/genetics/labo/Software/PermutCpSSR/index.html>). To test for the existence of phylogeographical structure, the comparison of G_{ST} versus N_{ST} was conducted using Permut based on 1,000 random permutations of haplotypes across populations. If N_{ST} is significantly higher than G_{ST} , then genealogically closely related haplotypes tend to occur together within populations, indicating the presence of phylogeographical structure [51]. Significance of P values was examined by the U Test using the Haplonst procedure [51].

To estimate the divergence times of cpDNA haplotypes, the BEAST version 1.6.1 [52] was carried out for lineage analysis, following a Bayesian relaxed molecular clock. A Bayesian Markov Chain Monte Carlo (MCMC) procedure was run for 20,000,000 generations with a coalescent-based tree prior rule and a HKY substitution model. Given the uncertainty of the cpDNA mutation rate, the normal distribution prior with a mean value of $2 \times 10^{-9} \text{ s}^{-1} \text{ yr}^{-1}$ (substitutions per site per year) and a SD of 6.08×10^{-9} for most angiosperm species were adopted as a criterion for the analytical operation [24, 53], to cover the rate variation within the 95% range of the distribution for the present calculation of molecular divergence times of genotypes. The combined parameters were validated in Tracer version 1.5 [52] to check whether the parameter value was greater than 1 and the effective sample size value was larger than 200, as an indication that these sequence data were suitable for a relaxed molecular clock model. Trees were finally compiled in FigTree 1.3.1. The statistical support of clades depended on Bayesian posterior probability.

To investigate spatial genetic patterns of geographical distance among populations for each defined regional group, genetic landscape shape analysis was performed using Alleles In Space [54]. Primarily, a connective Delaunay triangulation network was constructed based on the geographic coordinates of all sampled points, according to the Watson (1992) and Brouns (2003) methods [55, 56]. Secondly, the average genetic distances resulting from the cpDNA sequence data were calculated among populations in the network. Afterwards, the genetic distance matrix was interpolated onto spatial grids of the connective network, forming a landscape

shape. A final three-dimensional surface plot was generated, where the abscissa and ordinate signified geographical coordinates and the vertical coordinate were equivalent to genetic distance.

Possible demographic historical expansions for defined groups were inferred by mismatch distribution analysis using Arlequin with 1,000 parametric bootstrap replicates [57]. Populations that have experienced recent expansion are expected to have a unimodal curve in the pairwise mismatch distribution, whereas stable populations should have a bi- or multimodal shape [58, 59]. Neutrality tests (Tajima's D and Fu's F_s statistics) [60, 61] were used to detect further evidence for probable historical processes [62, 63]. A positive or negative D value (not 0) may imply that a population underwent expansion or bottlenecks [64], while a negative F_s value probably indicates a recent population expansion [61].

To explore the potential distribution changes of the species between the present-day and the Last Glacial Maximum (LGM; approximately 18–21 ka), the ecological niche modeling was performed in MAXENT version 3.3.1 [65], following the maximum entropy algorithm. We input an available coordinate set of occurrence points from the species' entire geographical distribution in Eastern Central Asia based on the Chinese Virtual Herbarium (<http://www.cvh.org.cn/cms/>) and our field survey locations (Table 1). The current climatic envelopes were predicted based on WorldClim dataset [66] and the LGM layers in accordance with the Model for Interdisciplinary Research on Climate (MIROC) [67] and the Community Climate System Model (CCSM) [68]. We acquired 19 BIOCLIM variables to model ecological niche at 2.5 arcmin resolution from the WorldClim database (available at <http://www.worldclim.org/download/>). To avoid the potential problems of overfitting [69], we removed the highly correlated bioclimatic variables, retaining nine least correlated ones. Model performance was identified by the area under the receiver operating characteristic curve (AUC). The default parameters to form ecological niche modeling were chosen as follows: regularization multiplier = 1.0, percentage of the data set for train the model = 75%, percentage for testing = 25%, and the threshold for available presence data = 10 percentiles. In addition, a Jackknife test was used to measure the significant contributions of the BIOCLIM variables on the occurrence prediction for each model. Finally, potential distribution ranges for the present-day and LGM were projected in ArcMap 9.3 (ESRI, Redlands, CA, USA).

3. Results

3.1. Chloroplast Variation and Haplotype Geographical Distribution. The aligned sequence length of the three cpDNA regions was 2254 bp (923 bp for *trnL-trnF*, 728 bp for *rps12-rpl20*, and 603 bp for *ndhF* (329F, 927R)). Based on the total sequences, thirty-six polymorphic sites (thirty nucleotide substitutions and six indels) were obtained (Table S2). A total of 24 haplotypes (H1–H24) were identified from 267 samples among 31 populations across the entire geographic range of *C. spinosa* (Figure 1(a), Table 1).

According to the results of SAMOVA, the 31 studied populations were divided into five defined phylogeographical

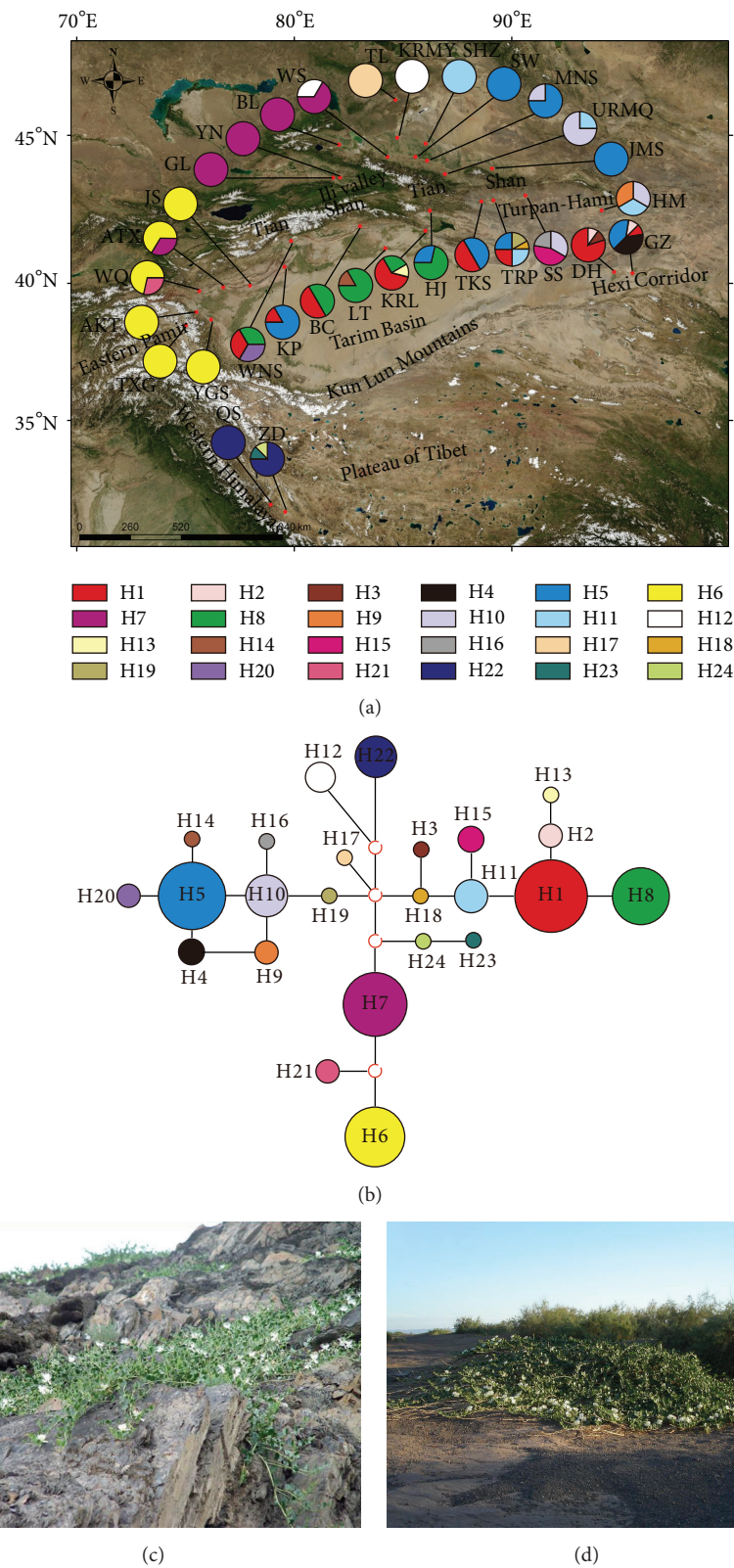


FIGURE 1: Geographical distribution and phylogenetic network of haplotypes in sampled locations of *Capparis spinosa*. (a) Distribution of the cpDNA haplotypes. The abbreviated letters of population localities correspond to those shown in Table 1. Pie charts show the different haplotypes and their proportions in each population; different colors are consistent with matching haplotypes in the figure legend. (b) Median-joining network reflecting the cpDNA haplotype relationships. Small, open red circles represent potentially intermediate genotypes with other mutational steps between real haplotypes; circle sizes are proportional to haplotype frequencies. (c, d) Two different kinds of habitats of the species: (c) eroding mountain slopes and (d) gravel deserts.

groups: (I) the Western Himalayas Group, (II) the Eastern Pamir Group, (III) the Ili Valley Group, (IV) the north side of the Tianshan Mountains Group, and (V) the south side of the Tianshan Mountains Group (Table 1). The main genealogical relationships of the network connection diagram (Figure 1(b)) as well as the BEAST tree (Figure 2) suggested the similar divergence trend.

Distribution of the haplotypes showed high geographic structure. Haplotypes H22, H6, H7, H1, H8, H5, and H10 had the most frequent and widespread distributions: H22 only was distributed in the populations of Group I; H6 occurred in all populations of Group II; H7 only occurred in the populations of Group III; H1 and H8 were dominantly distributed in the populations of Group V, while H5 and H10 were mainly present in the populations of Group IV (Figure 1(a)).

3.2. Genetic Diversity and Population Genetic Structure. Genetic diversity analysis revealed that total genetic diversity ($H_T = 0.916$, s.e. = 0.0172) was much higher than average gene diversity ($H_S = 0.316$, s.e. = 0.0533), which indicated that there was considerable population differentiation across the distributional range. The U Test showed that total N_{ST} (0.843, s.e. = 0.033) was significantly higher than G_{ST} (0.655, s.e. = 0.057) ($U = 6.91$, $P < 0.01$), indicating cpDNA variation with signal phylogeographic structure for the species.

AMOVA revealed that a larger proportion of the variation was interpopulational than intrapopulational (85.20% versus 14.80%). Differences among geographical groups were significant, and AMOVA results showed that 71.83% ($F_{CT} = 0.71831$, $P < 0.001$) of total variation existed between the five groups (Table 2). For the two genetic diversity indexes, haplotype diversities (H_d) within the 31 populations ranged from 0 to 0.8485, and nucleotide diversities (π) ranged from 0 to 0.0021. Considering the whole species, high degrees of haplotype diversity ($H_d = 0.8925$) and nucleotide diversity ($\pi = 0.0037$) were found. At the level of population, six populations (TRP, HM, WNS, GZ, SS, and KRL) in Group V showed considerably high degrees of genetic diversity (Table 1).

3.3. Molecular Divergence Time Estimates and Demographic History Analyses. The phylogenetic tree yielded by BEAST analysis was strongly coupled with geography, demonstrating that the coalescent sequences were suitable to be employed in phylogenetic and phylogeographic analysis. The results indicate that the identified cpDNA haplotypes are clustered into six main clades. Divergence times estimated from BEAST analysis were Pleistocene, and the main divergence between the geographical groups began at 1.179 Ma, in early Pleistocene (Figure 2).

The three-dimensional surface plot produced by genetic landscape shape analyses showed that spatial genetic distances gradually increased from the western populations to eastern populations (Figure 3). For mismatch analysis, the observed mismatch curve was not unimodal but strongly discordant with a model of sudden range expansion (Figure 4), indicating the stable populations. Additionally, no supporting evidence was provided by Tajima's D and Fu's F_s values for recent range expansion events.

3.4. Potential Distribution Modeling. Under the selected model, the test AUC value for the ENM of *C. spinosa* in the present-day was 0.995; and, under the MIROC climate scenario, the value was 0.997 in the LGM projection. SDM did not yield suitable or credible distribution in LGM based on the CCSM model. The Jackknife analysis of regularized training gain indicated that there were five environmental variables contributing more highly to the distribution model in the present-day: precipitation of the wettest month (22.64%), mean temperature of the coldest quarter (22.97%), precipitation of the coldest quarter (15.48%), mean temperature of the driest quarter (8.25%), and precipitation seasonality (6.37%). The five most contributive bioclimatic variables of the MIROC model in the LGM were annual mean temperature (24.32%), precipitation of the coldest quarter (20.18%), mean temperature of the coldest quarter (13.87%), mean temperature of the driest quarter (12.18%), and precipitation of the wettest month (11.89%). The range differences under the distinct climatic conditions of the two models revealed a diminished amount of suitable habitat for the species during the LGM period. In comparison with the present potential distribution, the MIROC scenario indicated that the high-altitude ranges of the species in the Western Himalayas, eastern Pamir, and western Tianshan Mountains were diminished (Figure 5), according to the LGM modeling distribution. However, some localities in the middle Tianshan, such as WNS and KRL, were not affected by this in the LGM model (Figure 5), and thus midelevation gorges environments might have provided suitable survival habitats for the species in the glacial epoch.

4. Discussion

4.1. Genetic Variation. At the species level, a high total gene diversity ($H_T = 0.916$) was detected. Compared with other species where three intergenic spacers have been sequenced, gene diversity in *C. spinosa* was higher, for instance, *Juniperus sabina* ($H_T = 0.577$) [70] and *Gymnocarpus przewalskii* ($H_T = 0.903$) [29]. The results of genetic diversity analysis also give indication of a high level of total haplotype diversity ($H_d = 0.8925$) (Table 1). High values of the diversity indices might have been promoted by the accumulation of variation in *C. spinosa*, which, to a great extent, may have been associated with differences in geological and topographical conditions, complex habitat patterns, and gradually intensive aridification occurring in Eastern Central Asia during the Quaternary period.

Dramatic genetic variation among the total populations was shown, accounting for a far greater proportion than that within populations (Table 2). The genetic differences among populations are believed to be closely related to the degree of gene exchange [71], which depends on seed transfer for cpDNA in most angiosperms [72]. The seed dispersal modes of *C. spinosa* mainly include anemochory, myrmecochory, and endozoochory by rodent vector [73]. In the current survey regions, the mountains and sprawling desert zones isolated habitat into disjunctive geographical units. In this situation, gene flow between populations was potentially restricted in short distance. Especially in

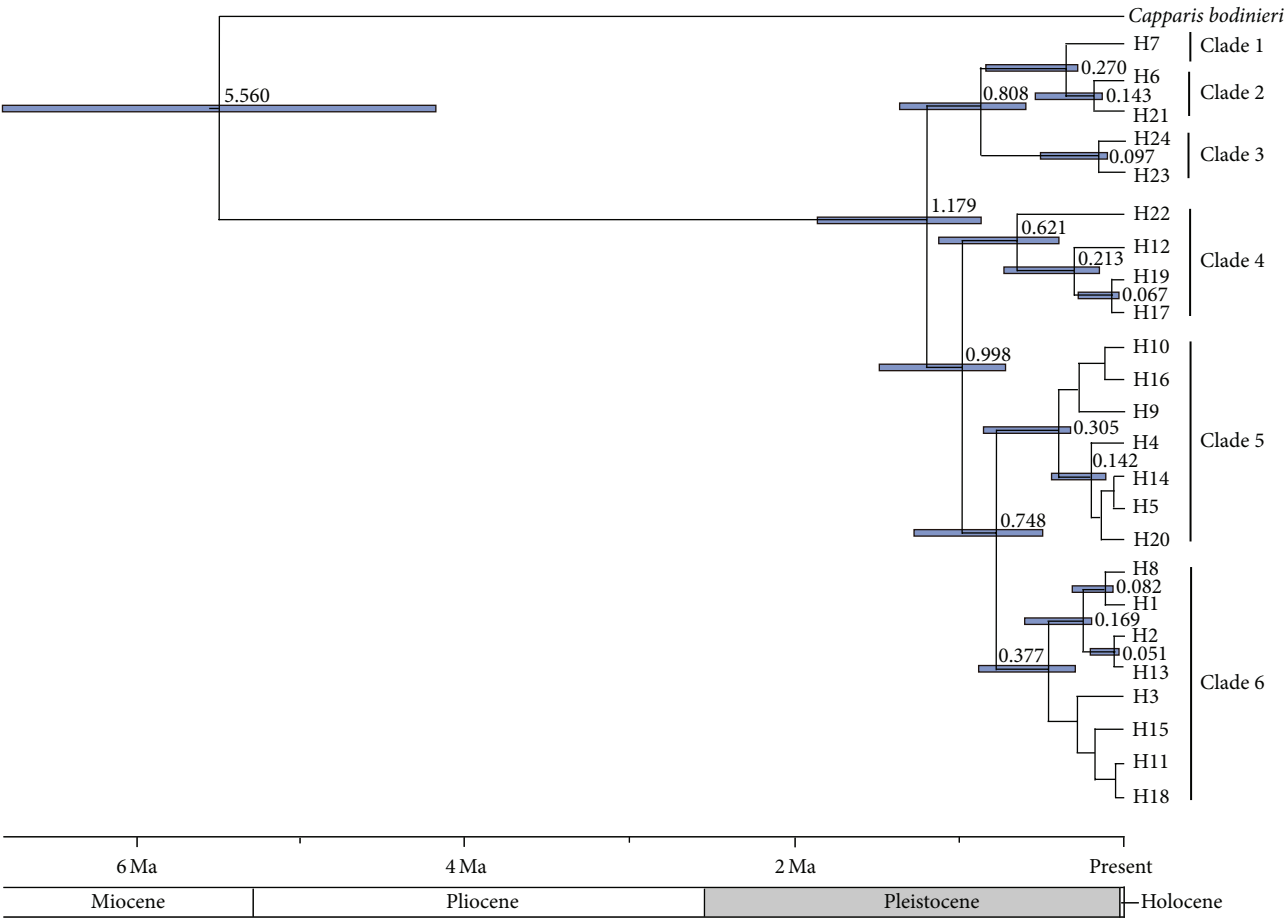


FIGURE 2: Phylogenetic chronogram of 24 haplotypes in *C. spinosa* generated from BEAST analysis. Divergence time (million years ago) of nodes are shown; blue horizontal bars indicating the 95% ranges of highest posterior density.

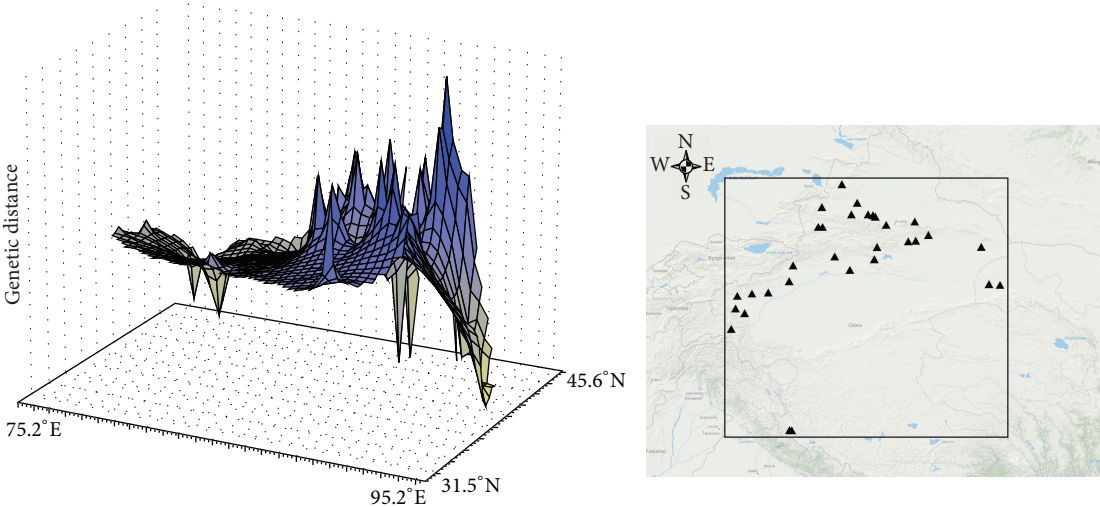
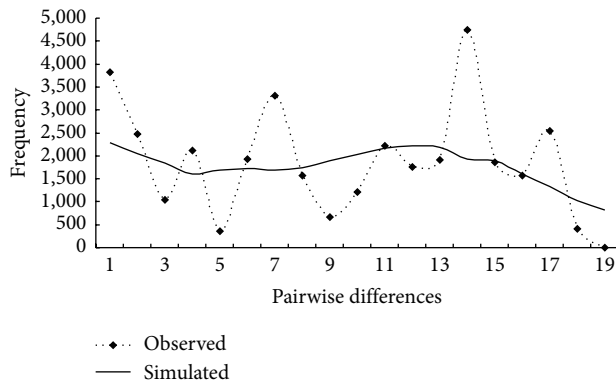


FIGURE 3: Spatial genetic landscape shapes constructed by interpolation analysis for *C. spinosa*. The abscissae and ordinates correspond to geographical coordinates covering the entire distributional populations, and the vertical axes represent genetic distances.

TABLE 2: Results of AMOVA analysis of molecular variance for populations and geographical groups of *C. spinosa*.

| Source of variation | d.f. | SS | VC | PV (%) | Fixation index |
|---------------------------------|------|----------|-------|--------|----------------------|
| Among groups | 4 | 738.961 | 3.723 | 71.83 | F_{CT} : 0.71831** |
| Among populations within groups | 26 | 175.341 | 0.693 | 13.37 | F_{SC} : 0.47458** |
| Within populations | 236 | 181.039 | 0.767 | 14.80 | F_{ST} : 0.85199** |
| Total | 266 | 1095.341 | 5.183 | | |

d.f.: degrees of freedom; SS: sum of squares; VC: variance components; PV: percentage of variation; F_{SC} : correlation of chlorotypes within populations relative to groups; F_{ST} : correlation within populations relative to the total; F_{CT} : correlation within groups relative to the total; ** $P < 0.001$; 1000 permutations.

FIGURE 4: Pairwise mismatch distributions of cpDNA sequences of *C. spinosa*.

small sized populations, individuals might be susceptible to inbreeding. Therefore, gene flow was limited attributing to the incapacity of seed dispersal across impassable physical barriers by anemochory or zoochory, resulting in population differentiation.

4.2. Influence of Pleistocene Geological and Climatic Changes on Phylogeographical Structure. The timescales of intraspecific divergences are usually related to historical geologic and climatic events [3, 12]. The estimated genetic divergence time between the five geographical groups started at approximately 1.179 million years ago based on molecular dating results of the BEAST analysis (Figure 2), which falls into early Pleistocene, the phase of intense uplift of Himalayas [74]. From early to middle Pleistocene, most large-scale piedmont glaciers developed on the northern slope of the Himalayas [75], which may have been the impetus causing haplotypes H23, H24, and H22 belonging to the Western Himalayan populations to branch off the phylogenetic tree (0.808–0.621 Ma) (Figure 2). At the intersection of the Himalaya, Kunlun, and Tianshan Mountains, the Pamir uplift was structured from at least the end of the Paleogene. In the Quaternary stage, owing to at least three large-scale glaciations, piedmont moraines intermittently formed in this region [76]. In connection with these fluctuations, the Western Himalayan and Eastern Pamirian populations from high-altitude localities experienced large-scale differentiation, accompanied with adaptation to colder climates, resulting in different degrees of genetic diversity and conspicuous population differences. We found that haplotypes H22 and H6 were, respectively, distributed over every sampled location in Western Himalayan

and Eastern Pamir region, while the rare haplotypes H23, H24, and H21 were randomly distributed in single populations. Although BEAST analysis and network diagram did not clearly show that haplotype H22 clustered into the same clade with H23 and H24, it occupies a key transitional position (Figures 1(b) and 2). Genetically H22 contained a haplotype approximating those in the Tianshan region, while geographically it is distributed in the Western Himalayan region, indicating a possible common Tethyan origin and then allopatric divergence and local evolution.

In comparison with the formation of the Himalayas and Pamirs, neotectonic movement of Tianshan Mountains was somewhat lagging, occurring in Pliocene and intensively uplifting in early Pleistocene [7, 8]. Since Quaternary diastrophism and climatic changes, together with more than three glacial-interglacial cycles [77], not only changed the topography but also resulted in temperature and precipitation differences between western and eastern ranges, north and south slopes, and high and low altitudes [74], rapid uplift of the Tianshan range as a barrier would, of necessity, have changed the local atmospheric circulation and screened the sources of moist airflow, intensifying the arid climate in adjacent regions. Later, since the mid-Pleistocene, enhanced drought and cold occurred in the Tianshan and the surroundings at 0.6–0.2 Ma [78], as evidenced by loess sediments on northern slope of the Tianshan Mountains [74]. During this period, a weakened southwest monsoon and strengthened plateau winter monsoon resulted in increasing aridification and desert expansion [79]. Compared with the geographic patterns under present-day climatic conditions, distribution ranges of *C. spinosa* during the LGM period were contracted (Figure 5). Although distribution areas of the species changed, the main distribution in the low-altitude eastern desert region was stable during glacial periods and only marginally influenced. The great spatial and genetic distances among locations in the eastern Tianshan and Hexi Corridor can be perceived in the 3D surface plots (Figure 3). This characteristic of the spatial genetic landscape analysis also demonstrates that the species has had relatively stable habitats in the low-altitude desert region during an extensive evolutionary history. At the stage of the mid-Pleistocene period, the continuously expanding Taklimakan [80] and intensive aridification accelerated the worsening drought and heat for the lower elevations of the eastern Tianshan, Turpan-Hami Basin, and Hexi Corridor populations, which must have influenced recent divergence among these areas. In general, genetic divergence among geographic groups in Pleistocene appears to have been driven mainly by intense uplift of the

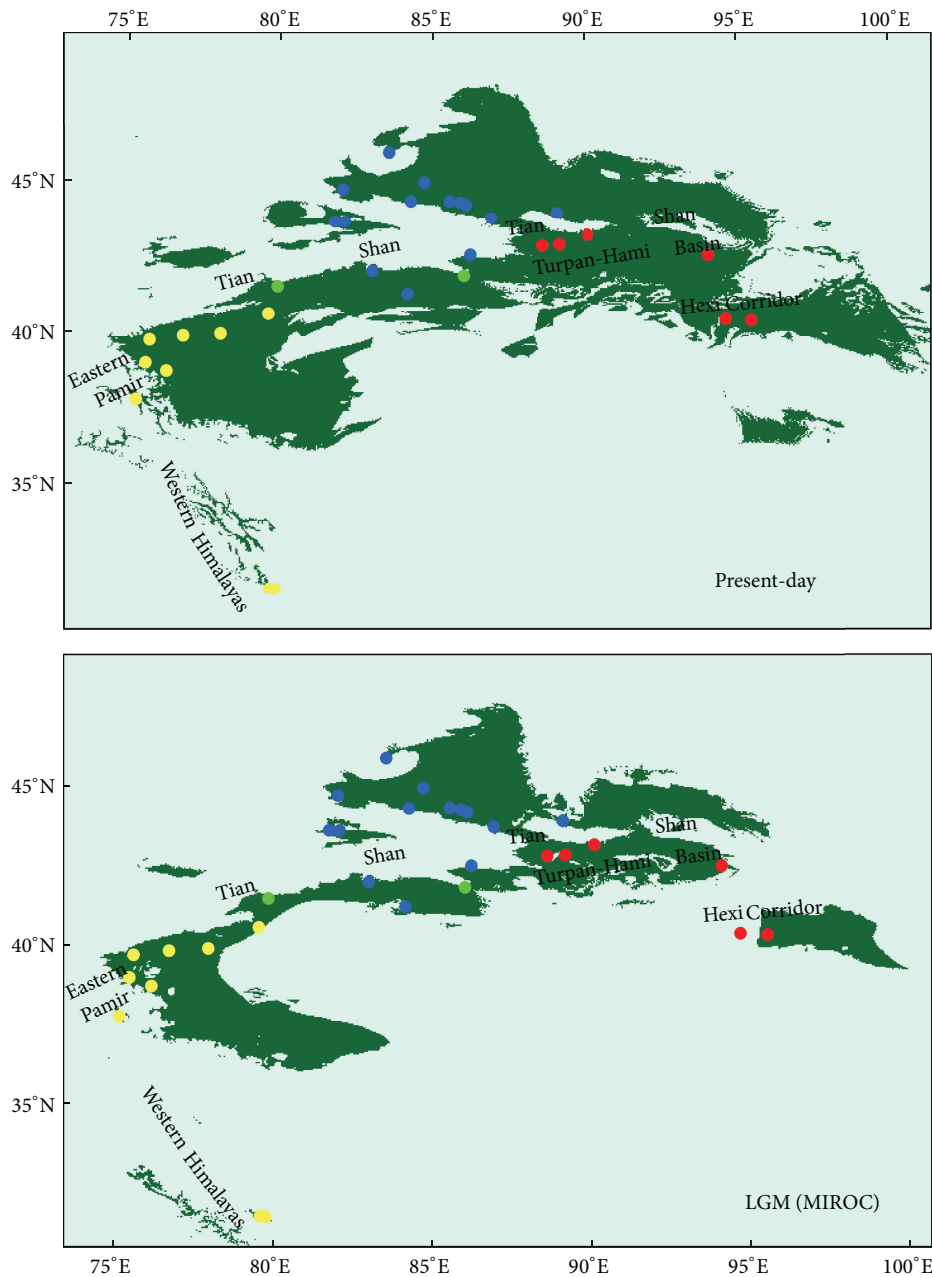


FIGURE 5: Species distribution modeling of *C. spinosa* during present-day and LGM periods in arid Eastern Central Asia. Potential distribution in the LGM epoch is based on scenarios of the MIROC model. Yellow localities: populations sensitive to glaciation in the high mountains. Blue localities: populations at middle elevations. Green localities: gorges served as refugia in the mid-Tianshan. Red localities: populations of low-altitude desert region.

Himalayas and Tianshan Mountains. The cold-dry to warm-humid climatic cycles during the late Quaternary are inferred to have promoted genetic divergence within groups.

4.3. Potential Glacial Refugia and Demographic Dispersal in Past Scenarios. According to previous glaciological research, maximum glaciation was likely to have occurred in the early phase of the mid-Pleistocene (0.8–0.6 Ma) in Eastern Central Asia [74, 77]. During this period and the next two glacial stages, owing to large-scale glacial advances, glaciers

in Tianshan at one time extended as far down as the piedmont belts [74]. In comparison with the present distribution range, the LGM localities of *C. spinosa* were reduced in the high-altitude ranges of the Western Himalayas, Pamirs, and western Tianshan Mountains (Figure 5), as a result of frigid climate in the wake of glacial advances. The increasingly severe climate in the glacial epoch would have damaged the ecological habitats of some xeric plant species [81]; however, it seems likely that *C. spinosa* could have survived in some stable gorges of Tianshan Mountains. These suitable

refugia were less influenced by climatic oscillations and therefore could maintain richer degrees of genetic diversity among populations [25, 82]. Refugial zones are generally recognizable because of high levels of genetic diversity and unique genotypes in species populations [12, 22] and are usually located in the middle altitude of mountains, which could supply warmer habitats for species to resist cold glacial climates [83]. A number of glacial refugia have been revealed worldwide on the basis of molecular approaches [84–86]. However, refugia localities in the Tianshan and adjacent arid regions have remained rather cryptic, as there seems to be a lack of applicable studies [23]. As a representative xerophyte, *C. spinosa* selected relatively warm gorges (WNS and KRL) in the middle Tianshan Mountains as potential refugia during glacial epochs, in accordance with the high levels of genetic diversity within these populations.

Unlike the traditional northward expansions during warm interglacial periods [18, 87], in the current study, the dispersal of *C. spinosa* was profoundly affected not only by temperature variation but also by aridification in Eastern Central Asia; thus xeric species may well prefer relatively arid areas during the warm and moist interglacial stages. Considering the network diagram, genotypes H1 and H8 were found with high rates of incidence on the south side of Tianshan group, indicating that these two were predominant haplotypes in Group V (Figure 1). These characteristics imply that the species should have undergone large-scale interglacial expansions from glacial refugia in this region. However, multimodal mismatch distribution shape (Figure 4) and positive Fu's F_s value failed to provide any significant evidence for recent range expansion. According to Printzen et al. (2003), extensive intraspecific differentiation is probably connected with range fragmentation or long-distance dispersal [88]. Evidenced from our field observations, *C. spinosa* is a xeromorphic species that prefers to distribute itself in eroding rocky mountains and arid piedmont proluvial or alluvial gravelly fans (Figures 1(c) and 1(d)). In the end of early Pleistocene and the early mid-Pleistocene, intensified aridification and the formation of Taklimakan desert resulted in dry and rainless conditions in piedmont zones of Tianshan [74, 89]. However, in contrast with the cold climate in glacial phase, in the interglacial and postglacial period the climate became warm and humid. At this time, glacier-melt water increased the volume of river runoff on the edge of the Tarim Basin [74]. Thus, the movement of the wind and river runoff provided appropriate conditions for survival and seed dispersal of *C. spinosa*. These could harbor relatively stable habitats, preserving the species' existing genetic variation and genetic differences among populations. There is considerable evidence, such as increasing $\delta^{18}\text{O}$ values, palynological components, multiple alluvial sequences, and lake sediments, that reveals the alternation of glacials and interglacials occurring in the Tianshan Mountains during the Pleistocene [74, 90, 91]. This circularly fluctuant climate propelled demographic dispersal in the interglacial intervals. Interestingly, unlike most other species with decreased levels of genetic diversity related to desert expansion during interglacial and postglacial colonization [29, 30], as xeric Tethyan relic, *C. spinosa* showed successively increased levels of genetic diversity and population

sizes, especially when it contacted the more droughty and hot gravelly deserts in Turpan-Hami Basin and Hexi Corridor. One reasonable explanation for this phenomenon is the probability that the suitable warm temperate continental climate in eastern gravel deserts is similar to the ancient hot and dry subtropical summer climate of the Tethyan Flora, preserving the characteristics of the original climate and environment. The other corollary is that the desert region could play the role of a shelter influenced only marginally by glaciation.

5. Conclusions

The current study focuses on how complicated Pleistocene geological and cyclical climatological events influenced the phylogeographic structure and genetic divergence of *C. spinosa* in arid Eastern Central Asia. Strongly spatial phylogeographic patterns were documented in the species. Intense uplift of the Himalaya and Tianshan Mountains around the early Pleistocene, coupled with a cold-dry climate and consequent aridification during the stage of Quaternary glaciation in this survey region, played important roles both in triggering and in shaping the current phylogeographic structure of the species group. There were at least three glacial-interglacial cycles in the Himalayas, as well as the Pamir and Tianshan Mountains during the Pleistocene. During glacial epochs, potential refugia were inferred for the gorges of the middle Tianshan Mountains, while, in interglacials and the postglacial period, the species experienced dispersal. The phylogeography of *C. spinosa* in the present research provides the basis for future studies on how xerophilous plant species across both arid high-altitude rocky mountains (Figure 1(c)) and low-altitude gravel deserts (Figure 1(d)) responded in a stepwise fashion to Quaternary geologic and climatic events.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

The authors are grateful to Dr. Stewart C. Sanderson in the Shrub Sciences Laboratory, Rocky Mountain Research Station, US Department of Agriculture, Utah, USA, for his helpful suggestions and English improvement of the paper. This study was financially supported by China National Key Basic Research Program (2014CB954201), CAS Key Laboratory of Biogeography and Bioresource in Arid Land, and Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences.

References

- [1] C. B. Cox and P. D. Moore, *Biogeography: an Ecological and Evolutionary Approach*, Blackwell Science, Oxford, UK, 2000.
- [2] R. I. Milne and R. J. Abbott, "The origin and evolution of tertiary relict floras," *Advances in Botanical Research*, vol. 38, pp. 281–314, 2002.
- [3] G. M. Hewitt, "The genetic legacy of the quaternary ice ages," *Nature*, vol. 405, no. 6789, pp. 907–913, 2000.

- [4] G. M. Hewitt, "Genetic consequences of climatic oscillations in the Quaternary," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 359, no. 1442, pp. 183–195, 2004.
- [5] E. Palamarev, "Paleobotanical evidences of the Tertiary history and origin of the Mediterranean sclerophyll dendroflora," *Plant Systematics and Evolution*, vol. 162, no. 1, pp. 93–107, 1989.
- [6] H. Sun, "Tethys retreat and Himalayas-Hengduanshan Mountains uplift and their significance on the origin and development of the Sino-Himalayan elements and Alpine flora," *Acta Botanica Yunnanica*, vol. 24, no. 3, pp. 273–288, 2002 (Chinese).
- [7] J.-M. Sun, R.-X. Zhu, and J. Bowler, "Timing of the Tianshan Mountains uplift constrained by magnetostratigraphic analysis of molasse deposits," *Earth and Planetary Science Letters*, vol. 219, no. 3–4, pp. 239–253, 2004.
- [8] L.-N. Wang, J.-Q. Ji, D.-X. Sun et al., "The uplift history of south-western Tianshan-Implications from AFT analysis of detrital samples," *Chinese Journal of Geophysics*, vol. 53, no. 4, pp. 931–945, 2010 (Chinese).
- [9] A. Takhtajan, T. J. Crovello, and A. Cronquist, *Floristic Regions of the World*, University of California Press, Berkeley, Calif, USA, 1986.
- [10] Z.-Y. Wu, H. Sun, Z.-K. Zhou, D.-Z. Li, and H. Peng, *Floristics of Seed Plants from China*, Science Press, Beijing, China, 2010 (Chinese).
- [11] L. L. Anderson, F. S. Hu, D. M. Nelson, R. J. Petit, and K. N. Paige, "Ice-age endurance: DNA evidence of a white spruce refugium in Alaska," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 33, pp. 12447–12450, 2006.
- [12] R. J. Petit, I. Aguinalalde, J.-L. de Beaulieu et al., "Glacial refugia: hotspots but not melting pots of genetic diversity," *Science*, vol. 300, no. 5625, pp. 1563–1565, 2003.
- [13] J. C. Avise, "The history and purview of phylogeography: a personal reflection," *Molecular Ecology*, vol. 7, no. 4, pp. 371–379, 1998.
- [14] A. Terrab, P. Schönswetter, S. Talavera, E. Vela, and T. F. Stuessy, "Range-wide phylogeography of *Juniperus thurifera* L., a presumptive keystone species of western Mediterranean vegetation during cold stages of the Pleistocene," *Molecular Phylogenetics and Evolution*, vol. 48, no. 1, pp. 94–102, 2008.
- [15] M. J. Hickerson, B. C. Carstens, J. Cavender-Bares et al., "Phylogeography's past, present, and future: 10 years after Avise, 2000," *Molecular Phylogenetics and Evolution*, vol. 54, no. 1, pp. 291–301, 2010.
- [16] J.-Q. Liu, Y.-S. Sun, X.-J. Ge, L.-M. Gao, and Y.-X. Qiu, "Phylogeographic studies of plants in China: advances in the past and directions in the future," *Journal of Systematics and Evolution*, vol. 50, no. 4, pp. 267–275, 2012.
- [17] A. C. Newton, T. R. Allnutt, A. C. M. Gillies, A. J. Lowe, and R. A. Ennos, "Molecular phylogeography, intraspecific variation and the conservation of tree species," *Trends in Ecology and Evolution*, vol. 14, no. 4, pp. 140–145, 1999.
- [18] I. V. Bartish, J. W. Kadereit, and H. P. Comes, "Late Quaternary history of *Hippophaë rhamnoides* L. (Elaeagnaceae) inferred from chalcone synthase intron (*Chsi*) sequences and chloroplast DNA variation," *Molecular Ecology*, vol. 15, no. 13, pp. 4065–4083, 2006.
- [19] X.-J. Ge, C.-C. Hwang, Z.-H. Liu et al., "Conservation genetics and phylogeography of endangered and endemic shrub *Tetraena mongolica* (Zygophyllaceae) in Inner Mongolia, China," *BMC Genetics*, vol. 12, article 1, 2011.
- [20] Y.-Y. Liao, A. W. Gichira, Q.-F. Wang, and J.-M. Chen, "Molecular phylogeography of four endemic *Sagittaria* species (Alismataceae) in the Sino-Japanese Floristic Region of East Asia," *Botanical Journal of the Linnean Society*, vol. 180, no. 1, pp. 6–20, 2016.
- [21] P. Schönswetter, I. Stehlik, R. Holderegger, and A. Tribsch, "Molecular evidence for glacial refugia of mountain plants in the European Alps," *Molecular Ecology*, vol. 14, no. 11, pp. 3547–3555, 2005.
- [22] J. R. Stewart, A. M. Lister, I. Barnes, and L. Dalén, "Refugia revisited: individualistic responses of species in space and time," *Proceedings of the Royal Society B: Biological Sciences*, vol. 277, no. 1682, pp. 661–671, 2010.
- [23] Y.-X. Qiu, C.-X. Fu, and H. P. Comes, "Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora," *Molecular Phylogenetics and Evolution*, vol. 59, no. 1, pp. 225–244, 2011.
- [24] D.-R. Jia, R.-J. Abbott, T.-L. Liu, K.-S. Mao, I. V. Bartish, and J.-Q. Liu, "Out of the Qinghai-Tibet Plateau: evidence for the origin and dispersal of Eurasian temperate plants from a phylogeographic study of *Hippophaë rhamnoides* (Elaeagnaceae)," *New Phytologist*, vol. 194, no. 4, pp. 1123–1133, 2012.
- [25] H. P. Comes and J. W. Kadereit, "The effect of quaternary climatic changes on plant distribution and evolution," *Trends in Plant Science*, vol. 3, no. 11, pp. 432–438, 1998.
- [26] O. Bettin, C. Cornejo, P. J. Edwards, and R. Holderegger, "Phylogeography of the high alpine plant *Senecio halleri* (Asteraceae) in the European Alps: in situ glacial survival with postglacial stepwise dispersal into peripheral areas," *Molecular Ecology*, vol. 16, no. 12, pp. 2517–2524, 2007.
- [27] X.-J. Ge, Y.-C. Chiang, C.-H. Chou, and T.-Y. Chiang, "Nested clade analysis of *Dunnia sinensis* (Rubiaceae), a monotypic genus from China based on organelle DNA sequences," *Conservation Genetics*, vol. 3, no. 4, pp. 351–362, 2002.
- [28] Z. Su, M. Zhang, and S. C. Sanderson, "Chloroplast phylogeography of *Helianthemum songaricum* (Cistaceae) from north-western China: implications for preservation of genetic diversity," *Conservation Genetics*, vol. 12, no. 6, pp. 1525–1537, 2011.
- [29] S. M. Ma, M. L. Zhang, and S. C. Sanderson, "Phylogeography of the rare *Gymnocarpus przewalskii* (Caryophyllaceae): indications of multiple glacial refugia in north-western China," *Australian Journal of Botany*, vol. 60, no. 1, pp. 20–31, 2012.
- [30] H.-H. Meng and M.-L. Zhang, "Diversification of plant species in arid Northwest China: species-level phylogeographical history of *Lagochilus Bunge ex Benth* (Lamiaceae)," *Molecular Phylogenetics and Evolution*, vol. 68, no. 3, pp. 398–409, 2013.
- [31] M. Jacobs, "The genus *Capparis* (Capparaceae) from the Indus to the Pacific," *Blumea*, vol. 12, pp. 385–541, 1965.
- [32] M.-L. Zhang and G. C. Tucker, "Capparaceae," in *Flora of China*, Z.-Y. Wu and P. H. Raven, Eds., pp. 433–450, Science Press/Missouri Botanical Garden Press, Beijing, China, 2008.
- [33] Z.-X. An, "Capparis L," in *Flora Xinjiangensis*, Z.-M. Mao, Ed., pp. 35–36, Xinjiang Science, Technology and Hygiene Publishing House, Ürümqi, China, 1995.
- [34] M. Zohary, "The species of *Capparis* in the Mediterranean and the near eastern countries," *Bulletin of the Research Council of Israel*, vol. 8, pp. 49–64, 1960.
- [35] R. N. Highton and J. R. Akeroyd, "Variation in *Capparis spinosa* L. in Europe," *Flora Europaea*, vol. 106, no. 2, pp. 104–112, 1991.

- [36] T. Zhang and D.-Y. Tan, "An examination of the function of male flowers in an andromonoecious shrub *Capparis spinosa*," *Journal of Integrative Plant Biology*, vol. 51, no. 3, pp. 316–324, 2009.
- [37] E. Saadaoui, A. Guetat, N. Tlili, M. El Gazzah, and A. Khaldi, "Subspecific variability of Tunisian wild populations of *Capparis spinosa* L.," *Journal of Medicinal Plants Research*, vol. 5, no. 17, pp. 4339–4348, 2011.
- [38] H. Nosrati, M. A. H. Feizi, M. Mazinani, and A. R. Haghighi, "Effect of population size on genetic variation levels in *Capparis spinosa* (Capparaceae) detected by RAPDs," *EurAsian Journal of BioSciences*, vol. 6, pp. 70–75, 2012.
- [39] Ö. Özbek and A. Kara, "Genetic variation in natural populations of *Capparis* from Turkey, as revealed by RAPD analysis," *Plant Systematics and Evolution*, vol. 299, no. 10, pp. 1911–1933, 2013.
- [40] S. O. Rogers and A. J. Bendich, "Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues," *Plant Molecular Biology*, vol. 5, no. 2, pp. 69–76, 1985.
- [41] J. D. Thompson, T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins, "The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools," *Nucleic Acids Research*, vol. 25, no. 24, pp. 4876–4882, 1997.
- [42] P. Librado and J. Rozas, "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data," *Bioinformatics*, vol. 25, no. 11, pp. 1451–1452, 2009.
- [43] H.-J. Bandelt, P. Forster, and A. Röhl, "Median-joining networks for inferring intraspecific phylogenies," *Molecular Biology and Evolution*, vol. 16, no. 1, pp. 37–48, 1999.
- [44] I. Dupanloup, S. Schneider, and L. Excoffier, "A simulated annealing approach to define the genetic structure of populations," *Molecular Ecology*, vol. 11, no. 12, pp. 2571–2581, 2002.
- [45] J. Godbout, J. P. Jaramillo-Correa, J. Beaulieu, and J. Bousquet, "A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range North American conifer," *Molecular Ecology*, vol. 14, no. 11, pp. 3497–3512, 2005.
- [46] D. Magri, G. G. Vendramin, B. Comps et al., "A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences," *New Phytologist*, vol. 171, no. 1, pp. 199–221, 2006.
- [47] M. Heuertz, S. Fineschi, M. Anzidei et al., "Chloroplast DNA variation and postglacial recolonization of common ash (*Fraxinus excelsior* L.) in Europe," *Molecular Ecology*, vol. 13, no. 11, pp. 3437–3452, 2004.
- [48] T. Iwasaki, K. Aoki, A. Seo, and N. Murakami, "Comparative phylogeography of four component species of deciduous broad-leaved forests in Japan based on chloroplast DNA variation," *Journal of Plant Research*, vol. 125, no. 2, pp. 207–221, 2012.
- [49] M. Nei, *Molecular Evolutionary Genetics*, Columbia University Press, New York, NY, USA, 1987.
- [50] L. Excoffier, G. Laval, and S. Schneider, "Arlequin (version 3.0): an integrated software package for population genetics data analysis," *Evolutionary Bioinformatics*, vol. 1, pp. 47–50, 2005.
- [51] O. Pons and R. J. Petit, "Measuring and testing genetic differentiation with ordered versus unordered alleles," *Genetics*, vol. 144, no. 3, pp. 1237–1245, 1996.
- [52] A. J. Drummond and A. Rambaut, "BEAST: Bayesian evolutionary analysis by sampling trees," *BMC Evolutionary Biology*, vol. 7, no. 1, article 214, 2007.
- [53] K. H. Wolfe, W. H. Li, and P. M. Sharp, "Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 84, no. 24, pp. 9054–9058, 1987.
- [54] M. P. Miller, "Alleles In Space (AIS): computer software for the joint analysis of interindividual spatial and genetic information," *Journal of Heredity*, vol. 96, no. 6, pp. 722–724, 2005.
- [55] D. F. Watson, *Contouring: A Guide to the Analysis and Display of Spatial Data (with Programs on Diskette)*, Pergamon Press, Oxford, UK, 1992.
- [56] G. Brouns, A. De Wulf, and D. Constaes, "Delaunay triangulation algorithms useful for multibeam echosounding," *Journal of Surveying Engineering*, vol. 129, no. 2, pp. 79–84, 2003.
- [57] S. Schneider and L. Excoffier, "Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA," *Genetics*, vol. 152, no. 3, pp. 1079–1089, 1999.
- [58] M. Slatkin and R. R. Hudson, "Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations," *Genetics*, vol. 129, no. 2, pp. 555–562, 1991.
- [59] A. R. Rogers and H. Harpending, "Population growth makes waves in the distribution of pairwise genetic differences," *Molecular Biology and Evolution*, vol. 9, no. 3, pp. 552–569, 1992.
- [60] F. Tajima, "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism," *Genetics*, vol. 123, no. 3, pp. 585–595, 1989.
- [61] Y.-X. Fu, "Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection," *Genetics*, vol. 147, no. 2, pp. 915–925, 1997.
- [62] L. A. Hurtado, T. Erez, S. Castrezana, and T. A. Markow, "Contrasting population genetic patterns and evolutionary histories among sympatric Sonoran Desert cactophilic *Drosophila*," *Molecular Ecology*, vol. 13, no. 6, pp. 1365–1375, 2004.
- [63] C. I. Smith and B. D. Farrell, "Range expansions in the flightless longhorn cactus beetles, *Moneilema gigas* and *Moneilema armatum*, in response to Pleistocene climate changes," *Molecular Ecology*, vol. 14, no. 4, pp. 1025–1044, 2005.
- [64] F. Tajima, "The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites," *Genetics*, vol. 143, no. 3, pp. 1457–1465, 1996.
- [65] S. J. Phillips, R. P. Anderson, and R. E. Schapire, "Maximum entropy modeling of species geographic distributions," *Ecological Modelling*, vol. 190, no. 3–4, pp. 231–259, 2006.
- [66] R. J. Hijmans, S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis, "Very high resolution interpolated climate surfaces for global land areas," *International Journal of Climatology*, vol. 25, no. 15, pp. 1965–1978, 2005.
- [67] H. Hasumi and S. Emori, *K-1 Coupled GCM (MIROC) Description*, Center for Climate System Research, University of Tokyo, Tokyo, Japan, 2004.
- [68] W. D. Collins, C. M. Bitz, M. L. Blackmon et al., "The Community Climate System Model version 3 (CCSM3)," *Journal of Climate*, vol. 19, no. 11, pp. 2122–2143, 2006.
- [69] A. T. Peterson and Y. Nakazawa, "Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*," *Global Ecology and Biogeography*, vol. 17, no. 1, pp. 135–144, 2008.

- [70] Y.-P. Guo, R. Zhang, C.-Y. Chen, D.-W. Zhou, and J.-Q. Liu, "Allopatric divergence and regional range expansion of *Juniperus sabina* in China," *Journal of Systematics and Evolution*, vol. 48, no. 3, pp. 153–160, 2010.
- [71] M. Slatkin, "Gene flow and the geographic structure of natural populations," *Science*, vol. 236, no. 4803, pp. 787–792, 1987.
- [72] D. E. McCauley, "The use of chloroplast DNA polymorphism in studies of gene flow in plants," *Trends in Ecology & Evolution*, vol. 10, no. 5, pp. 198–202, 1995.
- [73] T. Zhang, *The Reproductive Ecology on Capparis spinosa L. (Capparaceae)*, Xinjiang Agricultural University, Urumqi, China, 2008 (Chinese).
- [74] Y.-F. Shi, Z.-J. Cui, and Z. Su, *The Quaternary Glaciations and Environmental Variations in China*, Hebei Science and Technology Press, Shijiazhuang, China, 2006 (Chinese).
- [75] S. C. Porter, "Quaternary glacial record in Swat Kohistan, West Pakistan," *Geological Society of America Bulletin*, vol. 81, no. 5, pp. 1421–1446, 1970.
- [76] Xinjiang Integrated Survey Group of Chinese Academy of Sciences, *Xinjiang Geomorphology*, Science Press, Beijing, China, 1978 (Chinese).
- [77] X. Xu, A. Kleidon, L. Miller, S. Wang, L. Wang, and G. Dong, "Late Quaternary glaciation in the Tianshan and implications for palaeoclimatic change: a review," *Boreas*, vol. 39, no. 2, pp. 215–232, 2010.
- [78] J.-J. Li, "The patterns of environmental changes since late Pleistocene in Northwestern China," *Quaternary Science Reviews*, vol. 3, pp. 197–204, 1990.
- [79] X. Wu, F. Wang, and Z.-S. An, "The phase and height of tibetan plateau rise in late cenozoic," in *Loess Quaternary Geology Global Changes*, D.-S. Liu and Z.-S. An, Eds., pp. 1–13, Scientific Publishing House, Beijing, China, 1992.
- [80] Z. Wu, "Approach to the genesis of the Taklamakan desert," *Acta Geographica Sinica*, vol. 36, no. 3, pp. 280–291, 1981 (Chinese).
- [81] Y.-X. Liu, "A study on origin and formation of the Chinese desert floras," *Acta Phytotaxonomica Sinica*, vol. 33, no. 2, pp. 131–143, 1995 (Chinese).
- [82] A. Widmer and C. Lexer, "Glacial refugia: sanctuaries for allelic richness, but not for gene diversity," *Trends in Ecology and Evolution*, vol. 16, no. 6, pp. 267–269, 2001.
- [83] Y.-J. Zhang, M. Stöck, P. Zhang, X.-L. Wang, H. Zhou, and L.-H. Qu, "Phylogeography of a widespread terrestrial vertebrate in a barely-studied Palearctic region: green toads (*Bufo viridis* subgroup) indicate glacial refugia in Eastern Central Asia," *Genetica*, vol. 134, no. 3, pp. 353–365, 2008.
- [84] R. J. Abbott, L. C. Smith, R. I. Milne, R. M. M. Crawford, K. Wolff, and J. Balfour, "Molecular analysis of plant migration and refugia in the Arctic," *Science*, vol. 289, no. 5483, pp. 1343–1346, 2000.
- [85] T. Lacourse, R. W. Mathewes, and D. W. Fedje, "Late-glacial vegetation dynamics of the Queen Charlotte Islands and adjacent continental shelf, British Columbia, Canada," *Palaeogeography, Palaeoclimatology, Palaeoecology*, vol. 226, no. 1-2, pp. 36–57, 2005.
- [86] L.-Y. Wang, R. J. Abbott, W. Zheng, P. Chen, Y. Wang, and J. Liu, "History and evolution of alpine plants endemic to the Qinghai-Tibetan Plateau: *Aconitum gymnanthum* (Ranunculaceae)," *Molecular Ecology*, vol. 18, no. 4, pp. 709–721, 2009.
- [87] G. M. Hewitt, "Quaternary phylogeography: the roots of hybrid zones," *Genetica*, vol. 139, no. 5, pp. 617–638, 2011.
- [88] C. Printzen, S. Ekman, and T. Tønsberg, "Phylogeography of *Cavernularia hultenii*: evidence of slow genetic drift in a widely disjunct lichen," *Molecular Ecology*, vol. 12, no. 6, pp. 1473–1486, 2003.
- [89] H.-B. Zheng, C. M. Powell, K. Butcher, and J. Cao, "Late Neogene loess deposition in southern Tarim Basin: tectonic and palaeoenvironmental implications," *Tectonophysics*, vol. 375, no. 1–4, pp. 49–59, 2003.
- [90] Q. Wen and Y. Shi, "The quaternary climo-environment changes in Chaiwopu basin of Xinjiang region," *Chinese Geographical Science*, vol. 3, no. 2, pp. 147–158, 1993.
- [91] H. Lu, D. W. Burbank, and Y. Li, "Alluvial sequence in the north piedmont of the Chinese Tian Shan over the past 550 kyr and its relationship to climate change," *Palaeogeography, Palaeoclimatology, Palaeoecology*, vol. 285, no. 3-4, pp. 343–353, 2010.

Research Article

Reconstructing the Phylogeny of *Capsosiphon fulvescens* (Ulotrichales, Chlorophyta) from Korea Based on *rbcL* and 18S rDNA Sequences

Sang-Mi Sun,¹ Seung Hwan Yang,² Kirill S. Golokhvast,³ Bao Le,⁴ and Gyuhwa Chung¹

¹Department of Biotechnology, Chonnam National University, Yeosu, Chonnam 59626, Republic of Korea

²Center for Nutraceutical and Pharmaceutical Materials, Myongji University, Yongin, Gyeonggi 17058, Republic of Korea

³Educational Scientific Center of Nanotechnology, Far Eastern Federal University, Vladivostok 690-950, Russia

⁴Department of Biomedical and Electronic Engineering, Chonnam National University, Yeosu, Chonnam 59626, Republic of Korea

Correspondence should be addressed to Seung Hwan Yang; ymichigan@mju.ac.kr and Gyuhwa Chung; chung@chonnam.ac.kr

Received 7 November 2015; Revised 29 February 2016; Accepted 21 March 2016

Academic Editor: Peter F. Stadler

Copyright © 2016 Sang-Mi Sun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Capsosiphon fulvescens is a filamentous green algae in the class Ulvophyceae. It has been consumed as food with unique flavor and soft texture to treat stomach disorders and hangovers, and its economic value justifies studying its nutritional and potential therapeutic effects. In contrast to these applications, only a few taxonomic studies have been conducted on *C. fulvescens*. In particular, classification and phylogenetic relationships of the *C. fulvescens* below the order level are controversial. To determine its phylogenetic position in the class, we used *rbcL* and 18S rDNA sequences as molecular markers to construct phylogenetic trees. The amplified *rbcL* and 18S rDNA sequences from 4 *C. fulvescens* isolates (Jindo, Jangheung, Wando, and Koheung, Korea) were used for phylogenetic analysis by employing three different phylogenetic methods: neighbor joining (NJ), maximum parsimony (MP), and maximum likelihood (ML). The *rbcL* phylogenetic tree showed that all taxa in the order Ulvales were clustered as a monophyletic group and resolved the phylogenetic position of *C. fulvescens* in the order Ulotrichales. The significance of our study is that the 18S rDNA phylogenetic tree shows the detailed taxonomic position of *C. fulvescens*. In our result, *C. fulvescens* is inferred as a member of Ulotrichaceae, along with *Urospora* and *Acrosiphonia*.

1. Introduction

Capsosiphon fulvescens (C. Agardh) Setchell and N. L. Gardner, filamentous chlorophycean seaweed, is found in the North Atlantic [1, 2] and in the Northern Pacific, including Korea [3] and Japan [4]. Its natural habitat is the upper intertidal regions of coastal sediments and rocky shores which it shares with the common edible seaweed *Ulva prolifera* Mueller. *Capsosiphon fulvescens*, a filamentous green algae, reproduces by biflagellated isogametes released from bisexual gametophytes [5]. This seaweed is known to be a contaminant in *Porphyra* cultivation [6]. However, in the Southwestern province of Korea, it has been consumed as food with unique flavor and soft texture to treat stomach disorders

and hangovers [7], and its economic value justifies studying its nutritional and potential therapeutic effects. Several physiological studies conducted *in vitro* and *in vivo* have suggested that extracts of *C. fulvescens* had an inhibitory effect on melanogenesis in B16 cells [8], induced apoptosis in AGS gastric cancer cells [9], and reduced cholesterol levels in hypercholesterolemic rats [10]. The potential economic interest in *C. fulvescens* could justify its large-scale cultivation in both the laboratory and the field [11].

In contrast to attention to its applications, only a few taxonomical studies have been conducted regarding *C. fulvescens*. In particular, classification and phylogenetic relationships of the *C. fulvescens* below the order level are controversial. The reasons for discrepancies among classification schemes

include disagreements regarding the evaluation of morphological characters. For example, *C. fulvescens* produces gametes and zoospores like those of *Ulothrix* and *Urospora* [12] while it was also considered to be closely related to *Monostroma* by Migita [5] because of discontinuous reproductive patches near the thallus apex as well as similarity of gametes. However, Chihara [4] considered that it was closely related to *Percuriaria* and *Ulva* because fronds produce zygotes that germinate directly without formation of a thick-walled zygote [4]. It is often difficult to identify sympleiomorphies characters because similar characters can be derived from convergent or parallel evolution of Ulotrichales.

Molecular systematics in seaweeds has progressed rapidly with the use of PCR coupled with sequencing methods. This molecular approach has been effective in addressing many phylogenetic questions that had not been solved using phenotypic characters. The gene for the large subunit of ribulose-bisphosphate carboxylase (*rbcL*) located in the chloroplast genome and the 18S rDNA in the nuclear genome have been extensively used for the inference of phylogenetic relationships at higher taxonomic levels because of their slow synonymous nucleotide substitution rates and strong functional constraints of *rbcL* sequence that reduced the evolutionary rate of nonsynonymous substitutions. The first report to mention the classification of *C. fulvescens* collected from North Atlantic with a molecular marker (cf. 18S rDNA) was presented by Hayden and Waaland [13]. They suggested that *C. fulvescens* was in the order Ulotrichales, which was consistent with Nagata's report [12]. In 2008, Hanic and Lindstrom [14] also used 18S rDNA sequence to prove that *C. fulvescens* and *Pseudothrix borealis* (entity formerly called *C. groenlandicus*) do not belong in the same genus *Capsosiphon*.

In this study, for the first time, *C. fulvescens* have been examined using *rbcL* gene and 18S rDNA sequences of newly collected material from different provinces in South Korea and these sequences have been employed to understand the phylogenetic position of Korean *Capsosiphon fulvescens* in the Ulvophyceae.

2. Materials and Methods

2.1. Sample Collection and Culture. *C. fulvescens* thalli were collected from five different seaweed farms located in Jindo, Koheung, Wando, and Janghuyn, South Korea, during December 2011 to February 2012. They were washed several times in clean cold seawater and kept on ice until being returned to the laboratory. Seawater samples were collected from discrete depths using 10 Niskin bottles arranged on a conductivity, temperature, and depth (CTD) rosette. The entire contents of the bottles were gravity filtered onto a 47 mm Poretics membrane filter (GE Osmonics, Fairfield, CT, USA), with a pore size of 5 μm , held within a Millipore Swinnex filter holder (Millipore, Bedford, MA, USA). The filtration time varied between 30 min and 2 h; if the Niskin bottle was not completely filtered at the end of 2 h, the filter was processed, noting the amount of seawater filtered. Freshly collected plants were grown in seawater filtered medium in glass culture vessels at 8°C under 30–50 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$, 14 : 10 h LD cycle. The species was identified

microscopically, manually separated from other algae, and washed with tap water and distilled water. Authentic standard compounds were purchased from Tokyo Kasei Kogyo Co., Ltd. (Japan), and Supelco Inc. (Bellefonte, USA).

2.2. DNA Extraction. Whole cultured biomass of *Capsosiphon fulvescens* was freeze-dried, and genomic DNA from freeze dried materials was extracted using a modified hexadecyltrimethylammonium bromide (CTAB) method [15], in which samples (~0.5 g) were ground in 1 mL CTAB using a sterile mortar and pestle. DNA extracts were cleaned with a Wizard PCR purification system (Promega, Madison, WI, USA), according to the manufacturer's instructions.

2.3. PCR Amplification and Sequencing. Double-stranded amplification of the 18S rDNA and *rbcL* regions was performed in a total volume of 50 μL using 1.0 μL of total genomic DNA (10–20 ng) template. The PCR amplifications were performed with 1 unit of *Taq* DNA Polymerase PCR Buffer (Invitrogen), 1.5 mM MgCl_2 , 200 μM dNTP mix, and 2.5 μmol of each primer (PTC-200, MJ Research, Waltham, MA, USA). The *rbcL* and 18S rDNA genes were amplified using published primers and other primers designed from an alignment of available *rbcL* and 18S sequences (Table 1) with the following conditions.

The reactions were conducted using an initial denaturation at 94°C for 3 min, followed by 35 cycles each of 1 min at 94°C, 1 min at 50°C, and 1 min at 72°C, followed by final extension for 10 min at 72°C. The PCR products were analyzed by 1.2% (w/v) agarose gel electrophoresis and purified with a Wizard PCR purification system (Promega, Madison, WI, USA). The cleaned PCR products were sequenced using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems Foster City, CA, USA) following the manufacturer's instructions. Sequences were generated in both the forward and reverse directions using an automatic sequencer ABI PRISM 310 Genetic Analyzer (Applied Biosystems Foster City, CA, USA) (see Supplementary Materials available online at <http://dx.doi.org/10.1155/2016/1462916>).

2.4. Phylogenetic Analyses. Assemblies of the newly created DNA sequences in this study were carried out using the DNASTAR program (DNASTAR, Inc., Madison, WI, USA). The coverage of sequences determined was 2x. Multiple sequence alignments were performed using the ClustalX 2.1 with default parameters and manually edited. Other sequences were obtained from GenBank using BLASTN search with the identified sequences as queries. Twenty-five *rbcL* sequences belonging to Ulvophyceae were used to construct a phylogenetic tree with 2 outgroup sequences. Thirty-three 18S rDNAs were used to construct phylogenetic tree. For phylogenetic analyses, 2 different methods were applied: neighbor joining (NJ) and maximum parsimony (MP) using software Mega v6 [18] with complete deletion of gaps. In MP analyses, nucleotide positions and character state changes were weighted equally after removing uninformative characters. ML analysis was carried out using RAxML v8

TABLE 1: PCR and sequencing primers used in the present study.

| Primer | Sequence | Target | Direction |
|---------------------|-------------------------------------|-------------|-----------|
| RH1 ^a | 5'-ATGTCACCACAAACAGAACTAAAGC-3' | <i>rbcL</i> | Forward |
| Rbc571 ^c | 5'-TGTTTACGAGGTGGTCTTGA-3' | <i>rbcL</i> | Forward |
| <i>rbcL</i> -LongF | 5'-ATTCCAAGGTCCTCCACACGG-3' | <i>rbcL</i> | Forward |
| Rbc590 ^c | 5'-TCAAGACCACCTCGTAAACA-3' | <i>rbcL</i> | Reverse |
| 1385r ^a | 5'-AATTCAAATTTAATTTCTTTCC-3' | <i>rbcL</i> | Reverse |
| <i>rbcL</i> -LongR | 5'-GCAGTCAATTCAGGACTCCATTTACAAGC-3' | <i>rbcL</i> | Reverse |
| ABI ^b | 5'-GGAGGATTAGGGTCCGATTCC-3' | 18S | Forward |
| 18S-F1 | 5'-TTCATTGATCAAGAACGAAAGYYGGG-3' | 18S | Forward |
| 18S-5'END-F2 | 5'-GTCATATGCTTGTCTCAAAGATTAAGCC-3' | 18S | Forward |
| 18S-3'END-F3 | 5'-GACGATTAGATACCGTCGTAGTCTCAAC-3' | 18S | Forward |
| 18S-R1 | 5'-GCAGGGACGTAATCAACGCGA-3' | 18S | Reverse |
| 18S-5'END-R1 | 5'-CCTTGTTACGACTTCTCCTTCCTCTAA-3' | 18S | Reverse |

^aManhart 1994 [16].^bvan Oppen 1995 [17].^cHayden and Waaland 2002 [13].

[19] with the HKY model of DNA sequence evolution and gaps were treated as unknown characters. For all 3 analysis types, branch support was assessed by bootstrapping (1000 replicates).

3. Results

3.1. Constitution of *rbcL* Phylogenetic Trees. *C. fulvenscens* *rbcL* sequence was aligned with 25 previously published *rbcL* sequences which represent Ulvales and Ulotrichales in class Ulvophyceae (Figure 1).

Sequences of *Myrmecia biatorellae* and *Chlorella vulgaris* in class Trebouxiophyceae were used as outgroups (Figure 1). For phylogenetic analyses, 3 different methods were applied: neighbor joining (NJ), maximum parsimony (MP), and maximum likelihood (ML). To construct the NJ tree, 1,253 nucleotide positions were included. Heuristic searches under the MP criterion with 339 parsimony informative characters recovered the 3 most parsimonious trees (tree length [L]: 545, consistency index [CI]: 0.454, and retention index [RI]: 0.645). The ML tree inferred with the HKY DNA substitution model also recovered a tree with the 3240.16 – ln L score. For clarity, only bootstrap numbers over 50% majority-rule consensus trees of MP and NP were shown on nodes in the ML tree. Within Ulotrichales, the clade comprising *Capsosiphon fulvenscens* and *Protomonostroma undulatum* was strongly supported as sister to the clade comprised of *Pseudothrix borealis* and the *Urospora* accessions.

3.2. Constitution of 18S rDNA Phylogenetic Tree. The *C. fulvenscens* 18S rDNA sequences were aligned with 33 previously published green algal sequences (Figure 2) at 1,439 nucleotide positions after removing all gaps.

Of the 33 sequences, 13 sequences represented Ulotrichaceae, Gomontiaceae, Gayraliaceae, and Monostromataceae

in Ultrichales and the remaining sequences represented Ulvellaceae, Kornmanniaceae, Bolbocoleaceae, Phaeophyllaceae, and Ulvaceae in Ulvales. In the 18S rDNA analysis, *Capsosiphon fulvenscens* occurred within a clade containing *Acrosiphonia*, *Urospora*, *Pseudothrix*, and *Protomonostroma*. It is closely related to *Protomonostroma* with strong support. HKY DNA substitution model was used to construct the ML tree with the score – ln L = 2022.00.

4. Discussion

After Hayden and Waaland [13] reported that Ulvales and Ulotrichales sensu Floyd and O'Kelly [20] are monophyletic sister orders, the systematics in Ulvales has been well supported by molecular marker as 18S rDNA [21, 22] and shows 3 main families: Ulvaceae, Kornmanniaceae, and Ulvellaceae. However, classification and phylogenetic relationships in Ulotrichales are complicated. Despite several studies of the beneficial value of *C. fulvenscens* to health, the position of *C. fulvenscens* in the order Ulotrichales is still unclear. Phylogenetic trees were constructed on the basis of *rbcL*, 18S rDNA, and combined sequences. This information has been further used for phylogenetic analysis and classification of Ulvophyceae and other related taxa.

Phylogenetic trees based on *rbcL* and 18S rDNA sequences exhibit topological differences that are due to the different rates at which these genes evolve. To make any statement from our phylogenetic trees, our study only selected the tree topology supported by >50% in at least 2 different phylogenetic trees. The *rbcL* tree (Figure 1) recovered Ulvales as a monophyletic group, which was consistent with the previous studies and showed that *C. fulvenscens* was weakly supported in Ulotrichales because all phylogenetic trees have showed a corresponding topology with at least 50% of bootstrap value.

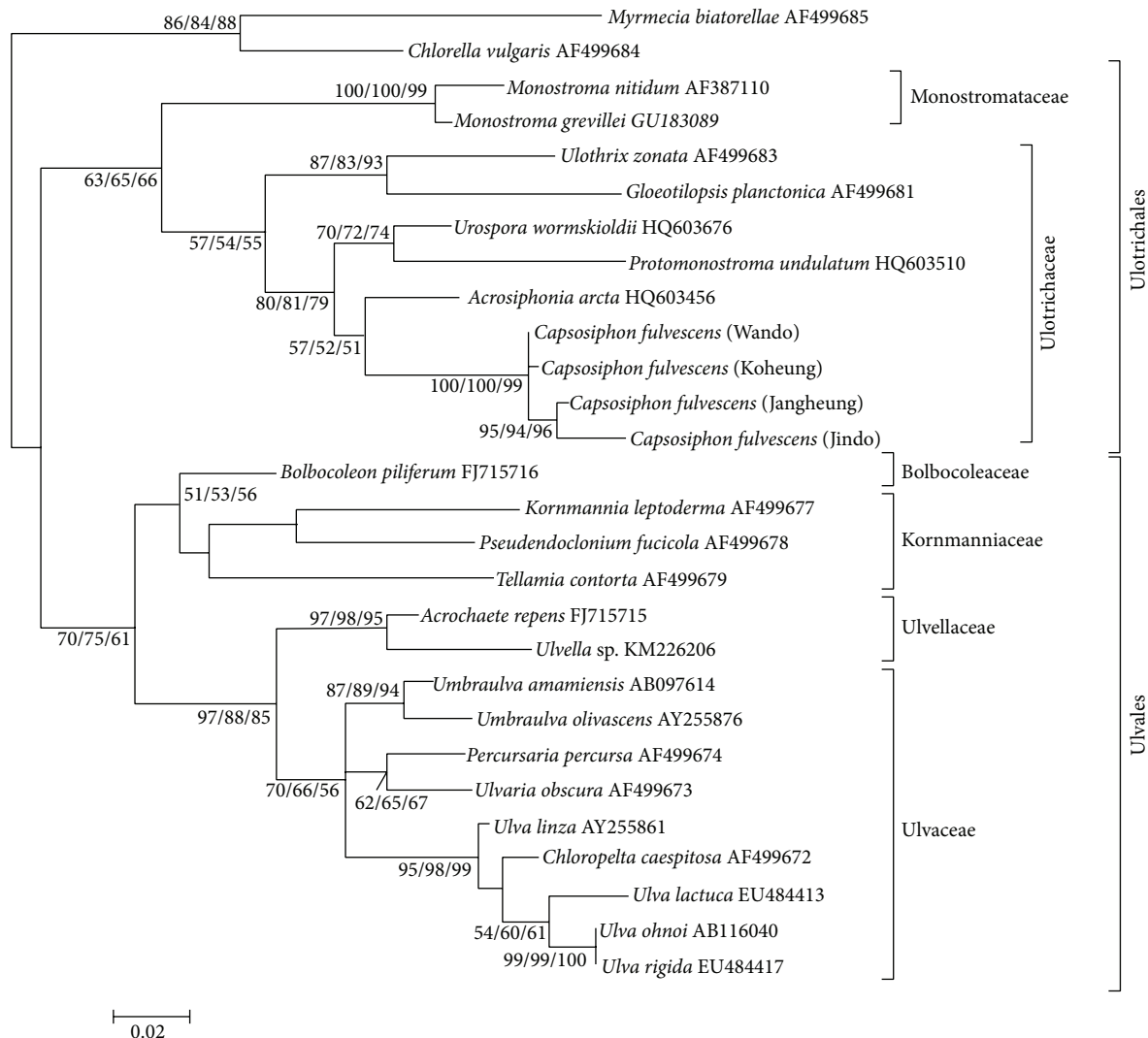


FIGURE 1: Maximum likelihood tree ($-\ln L = 3240.16$) of *rbcL* gene sequences of 26 species of Ulvophyceae. Numbers on branches represent maximum likelihood, neighbor joining, and maximum parsimony bootstrap support analysis, respectively (ML/NJ/MP). Scale bar: 0.02 (maximum composite likelihood).

The 18S rDNA phylogenetic tree (Figure 2) provided more information about the phylogenetic systematics of Ulotrichales and the position of *C. fulvenscens*. *Capsosiphon fulvenscens* appeared on its own branch of the Ulotrichales in the NJ and ML analysis; in the MP analysis, it appeared on its own branch at the base of the Ulotracheaceae. In all analyses, it clearly belonged to the Ulotrichales since the subtending branch, separating the Ulotrichales from members of the Ulvales, had 70% bootstrap support.

Our phylogenetic tree shows that Ulotrichales appears to be paraphyletic. Two clearly different lineages in Ulotracheaceae (*Acrosiphonia* and *Urospora*) were also successfully recovered with >80% of bootstrap numbers in all 3 phylogenetic methods (ML/MP/NJ) = 81/83/85, which was also consistent with the SSU rDNA tree [14, 21]. The key result here is that the tree topology in Ulotracheaceae shows 4 distinct lineages:

Gloeotilopsis, *Acrosiphonia*, *Capsosiphon*, and *Urospora* + *Protomonostroma* (Figure 2). This tree topology [*Acrosiphonia*, and (*Capsosiphon*, *Urospora* + *Protomonostroma*)] was supported by >50% of bootstraps in all 3 different phylogenetic trees, suggesting that *Capsosiphon* is more closely related to *Urospora* + *Protomonostroma*. In both MP and NJ analyses with the 18S rDNA data set, 3 well-supported groups were recovered: (1) Gomontiaceae, (2) Monostromataceae, and (3) Ulotracheaceae. The relationship between Ulotracheaceae and other genera is not well supported in either NJ or MP tree. In NJ tree, the proximal outgroup Ulotracheaceae and Gomontiaceae + Monostromataceae are similar to the phylogram including 5 representatives of Kornmanniaceae [22].

The present study showed that both phylogenetic analysis based on 18S rDNA and *rbcL* sequences analysis resolved

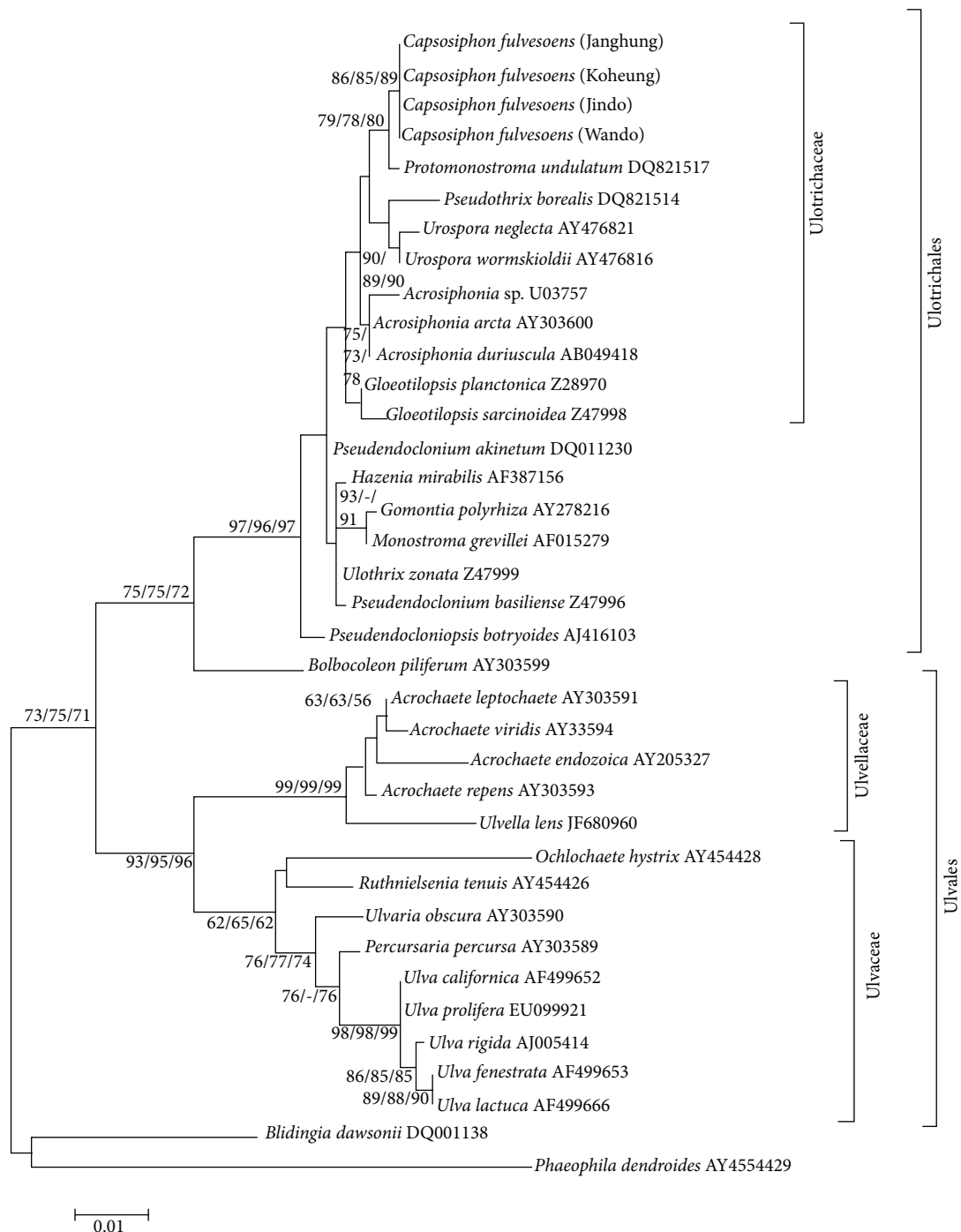


FIGURE 2: Maximum likelihood tree ($-\ln L = 2022.00$) of 18S rDNA gene sequences of 34 species of Ulvophyceae. Numbers on branches represent maximum likelihood, neighbor joining, and maximum parsimony bootstrap support analysis, respectively (ML/NJ/MP). Scale bar: 0.01 (maximum composite likelihood).

the phylogenetic position of *C. fulvescens* in Ulotrichales (Figures 1 and 2). All of 4 *C. fulvescens* were clustered with *Urospora wormskioldii* and *Acrosiphonia arcta* in the family Ulotrichaceae, Ulotrichales, with high bootstrap number in all phylogenetic trees (Figure 2), suggesting that *Capsosiphon* is a genus in the order Ulotrichales, although additional information is needed to support this hypothesis.

We will further study the vegetative morphology of this genus to compare it with that of known genera in the Ulotrichales.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

Seung Hwan Yang and Gyuhwa Chung contributed equally to this work.

Acknowledgments

The project was supported by the Russian Fund of Basic Research (15-04-02979).

References

- [1] C. Bliding, "A critical survey of European taxa in Ulvales. Part I. Capsosiphon, Percusaria, Blidingia, Enteromorpha," *Opera Botanica*, vol. 8, no. 3, pp. 1–160, 1963.
- [2] A. C. Mathieson and E. J. Hehre, "A comparison of the marine algae from the Goleta Slough and adjacent open coast of Goleta/Santa Barbara, California with those in the southern Gulf of Maine," *Rhodora*, vol. 96, pp. 207–258, 1994.
- [3] J. W. Kang, "On the geographical distribution of marine algae in Korea," *Bulletin of Pusan Fisheries College*, vol. 7, no. 1, pp. 1–125, 1966.
- [4] M. Chihara, "Developmental morphology and systematics of *Capsosiphon fulvescens* as found in Izu," *Japan. Bulletin of the National Science Museum, Tokyo*, vol. 10, pp. 163–170, 1967.
- [5] S. Migita, "Life cycle of *Capsosiphon fulvescens* (C. Agardh) Setchell and Gardner," *Bulletin of the Faculty of Fisheries, Nagasaki University*, vol. 22, pp. 21–31, 1967.
- [6] C. H. Sohn, "Porphyra, Undaria and Hizikia cultivation in Korea," *ALGAE*, vol. 8, no. 2, pp. 207–216, 1993.
- [7] C. H. Sohn, "The seaweed resources of Korea," in *Seaweed Resources of the World*, A. T. Critchley and M. Ohno, Eds., pp. 15–33, JICA, 1998.
- [8] Y. J. Mun, H. J. Yoo, H. E. Lee, J. H. Kim, H. B. Pyo, and W. H. Woo, "Inhibitory effect on the melanogenesis of *Capsosiphon fulvescens*," *Yakhak Hoeji*, vol. 49, pp. 375–379, 2005.
- [9] M.-J. Kwon and T.-J. Nam, "A polysaccharide of the marine alga *Capsosiphon fulvescens* induces apoptosis in AGS gastric cancer cells via an IGF-IR-mediated PI3K/Akt pathway," *Cell Biology International*, vol. 31, no. 8, pp. 768–775, 2007.
- [10] M.-J. Kwon and T.-J. Nam, "Effects of mesangi (*Capsosiphon fulvescens*) powder on lipid metabolism in high cholesterol fed rats," *Journal of the Korean Society of Food Science and Nutrition*, vol. 35, no. 5, pp. 530–535, 2006.
- [11] E. K. Hwang, Y. H. Yi, W. J. Shin, and C. H. Sohn, "Growth and maturation of a green alga, *Capsosiphon fulvescens*, as a new candidate for seaweed cultivation in Korea," in *Proceedings of the Seventeenth International Seaweed Symposium*, R. O. Chapman, R. J. Anderson, V. J. Vreeland, and I. R. Davison, Eds., pp. 59–64, Oxford University Press, 2003.
- [12] K. Nagata, "On the life history of *Urospora mirabilis* Areschoug from Muroan," *Bulletin of the Japanese Society for Phycology*, vol. 19, pp. 97–103, 1971.
- [13] H. S. Hayden and J. R. Waaland, "Phylogenetic systematics of the *Ulvaceae* (Ulvales, Ulvophyceae) using chloroplast and nuclear DNA sequences," *Journal of Phycology*, vol. 38, no. 6, pp. 1200–1212, 2002.
- [14] L. A. Hanic and S. C. Lindstrom, "Life history and systematic studies of *Pseudothrix borealis* gen. et sp. nov. (=North Pacific *Capsosiphon groenlandicus*, Ulotrichaceae, Chlorophyta)," *ALGAE*, vol. 23, no. 2, pp. 119–133, 2008.
- [15] J. J. Doyle and J. L. Doyle, "Isolation of plant DNA from fresh tissue," *Focus*, vol. 12, pp. 13–15, 1990.
- [16] J. R. Manhart, "Phylogenetic analysis of green plant *rbcl* sequences," *Molecular Phylogenetics and Evolution*, vol. 3, no. 2, pp. 114–127, 1994.
- [17] M. J. H. van Oppen, *Tracking trails by cracking codes [Ph.D. thesis]*, University of Groningen, Groningen, The Netherlands, 1995.
- [18] K. Tamura, G. Stecher, D. Peterson, A. Filipski, and S. Kumar, "MEGA6: molecular evolutionary genetics analysis version 6.0," *Molecular Biology and Evolution*, vol. 30, no. 12, pp. 2725–2729, 2013.
- [19] A. Stamatakis, "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies," *Bioinformatics*, vol. 30, no. 9, pp. 1312–1313, 2014.
- [20] G. L. Floyd and C. J. O'Kelly, "Phylum Chlorophyta, class Ulvophyceae," in *Handbook of Protoctista*, L. Margulis, J. O. Corliss, M. Melkonian, and D. J. Chapman, Eds., vol. 1, pp. 617–640, Jones & Bartlett, Boston, Mass, USA, 1990.
- [21] C. J. O'Kelly, W. K. Bellows, and B. Wysor, "Phylogenetic position of *Bolbocoleon piliferum* (Ulvophyceae, Chlorophyta): evidence from reproduction, zoospore and gamete ultrastructure, and small subunit rRNA gene sequences," *Journal of Phycology*, vol. 40, no. 1, pp. 209–222, 2004.
- [22] C. J. O'Kelly, B. Wysor, and W. K. Bellows, "Gene sequence diversity and the phylogenetic position of algae assigned to the genera *Phaeophila* and *Ochlochaete* (Ulvophyceae, Chlorophyta)," *Journal of Phycology*, vol. 40, no. 4, pp. 789–799, 2004.

Research Article

Molecular Typing Characteristic and Drug Susceptibility Analysis of *Mycobacterium tuberculosis* Isolates from Zigong, China

Hai-Can Liu,¹ Jian-Ping Deng,² Hai-Yan Dong,¹ Ti-Quan Xiao,² Xiu-Qin Zhao,¹ Zheng-Dong Zhang,² Yi Jiang,¹ Zhi-Guang Liu,¹ Qun Li,² and Kang-Lin Wan¹

¹State Key Laboratory for Infectious Diseases Prevention and Control, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China

²Zigong Center for Disease Control and Prevention, Zigong, Sichuan 643000, China

Correspondence should be addressed to Kang-Lin Wan; wankanglin@icdc.cn

Received 11 November 2015; Revised 5 January 2016; Accepted 14 January 2016

Academic Editor: Peter F. Stadler

Copyright © 2016 Hai-Can Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

China is one of the 22 countries with high TB burden worldwide, and Sichuan contained the second-largest number of TB cases among all of the Chinese provinces. But the characteristics of *Mycobacterium tuberculosis* circulated in Zigong, Sichuan, were still unknown. To investigate the character and drug resistance profile, 265 clinical isolates were cultured from tuberculosis patient's sputum samples in the year of 2010, of which the genetic profile was determined by using Spoligotyping and MIRU-VNTR typing methods, and the drug sensibility testing to the four first-line and four second-line antituberculosis (anti-TB) drugs was performed by using proportion method on Lowenstein-Jensen (L-J) media. The major Spoligotype was Beijing family (143/265, 53.96%), followed by T (80/265, 30.19%) and H (9/265, 3.40%) genotypes; the total Hunter-Gaston discrimination index (HGDI) of the 24 loci MIRU-VNTR was 0.9995. About 27.17% (72/265) of the isolates were resistant to at least one of the eight tested anti-TB drugs, and for Beijing and non-Beijing family isolates the proportion of drug resistance was 28.47% (41/144) and 25.62% (31/121), respectively. That is, the most prevalent genotype here was Beijing family, and the 24 loci VNTR analysis could supply a high resolution for genotyping, and Beijing and non-Beijing isolates had no difference ($p > 0.05$) for drug resistance.

1. Introduction

Tuberculosis (TB) remains an important problem for the public health worldwide, especially in developing countries. Of the five countries with the largest number of incident cases in 2012, China (0.9 million–1.1 million) ranked the second just next to India, which accounted for 12% of the global cases alone [1]. Recently the emerging of drug-resistant TB (DR-TB) is becoming an important threat to TB control and public health worldwide [2, 3], and according to the National Survey of Drug-Resistant Tuberculosis in China, 5.7% of new TB patients and 25.6% of the previously treated cases were multi-drug-resistant TB (MDR-TB) [4]. In the last decade, genotyping of *Mycobacterium tuberculosis* (*M. tuberculosis*) has significantly enhanced our understanding of

TB epidemiology by detecting suspected outbreaks and by tracing the infection routine. Such molecular methods like spacer oligonucleotide typing (Spoligotyping) and mycobacterial interspersed repetitive unit- (MIRU-) variable number tandem repeats (VNTR) typing have been used to characterize clinical *M. tuberculosis* isolates in China [5–7]. Obtaining the characteristic of *M. tuberculosis* isolates circulating in an area is helpful for understanding and controlling the spread of TB infection.

Although molecular epidemiology study of TB has been taken out in some area in China [8–10], few efforts have been made for the collection of DNA fingerprinting data based on the mutation profiles of isolates from the western region of China such as Sichuan province. Sichuan is a province that is located in southwestern China, with a population of

approximately 87 million inhabitants. Sichuan contained the second-largest number of TB cases among Chinese provinces, where the prevalence of both TB and drug-resistant TB was much higher than the average level in China [11]. The total cases of TB in Sichuan province are estimated to be about 272,000 in 2006 and the new cases found in 2006 were more than 70,000 [12]. So the genetic and drug-resistant characteristics analysis of clinical *M. tuberculosis* isolates from this area may give us some benefits on TB control in China.

The aim of this study was to identify and genetically characterize the *M. tuberculosis* isolates circulating in Zigong, southern Sichuan, and to explore the distribution of drug resistance profiles across the major Spoligotyping-defined *M. tuberculosis* genotypes.

2. Material and Methods

2.1. Bacterial Strains and DNA Samples. A total of 265 clinical *M. tuberculosis* isolates were collected from Zigong, Sichuan, in 2010. Chromosomal DNA was extracted from fresh bacteria cultures on the Lowenstein-Jensen (L-J) medium. A loopful of mycobacterial colonies were suspended in 400 μ L TE buffer (pH 8.0) and then boiled for 10 min. The suspension was centrifuged at 12,000 rpm for 10 min, and the supernatant was served as PCR template and stored at -20°C .

2.2. Spoligotyping. Spoligotyping was performed as described previously by Kamerbeek et al. [13]. The results were entered into Excel spreadsheets in a binary format and compared with SITVITWEB database [14], which is a publicly available international multimarker database for studying *Mycobacterium*.

2.3. MIRU-VNTR Typing. VNTR analysis was performed by the amplification of 24 genomic loci, the scheme of which has been described by Supply et al. [15]. Each VNTR locus was amplified individually in 20 μ L reaction mixture containing 1x PCR GC Buffer I (Takara Bio Inc.), 5 pmol of each primer set, 200 μ M of each dNTP, 0.5 U of DNA Taq Polymerase (Takara Bio Inc.), and 1.5 ng of DNA template. An initial denaturation step of 5 min at 95°C was followed by 35 cycles at 94°C for 1 min, 60°C for 1 min, and 72°C for 1 min, with a final extension at 72°C for 10 min. The PCR products were analyzed by electrophoresis on 2% agarose gel using 100 bp DNA ladder as size markers. The H37Rv strain was included as a positive control for each amplification reaction and run as an additional control for accuracy, and then the copy number of each locus was calculated by using the Image Lab software (Bio-Rad).

2.4. Drug Susceptibility Testing. Drug susceptibility testing of those 265 strains to the four first-line drugs (rifampicin [RFP], isoniazid [INH], streptomycin [SM], and ethambutol [EMB]) and four second-line drugs (amikacin [AMK], kanamycin [KA], ofloxacin [OFLX], and capreomycin [CPM]) was performed by using proportion method recommended by WHO [16]. Multi-drug-resistant TB (MDR-TB) is the isolate that was resistant to at least the two

first-line anti-TB drugs, RFP and INH, based on the WHO definition [17].

2.5. Computer Analysis. The Spoligotyping and MIRU-VNTR typing results of all the *M. tuberculosis* clinical isolates were analyzed by the BioNumerics program (Windows 7, version 5.10; Applied Maths, Kortrijk, Belgium). Clusters were defined as groups that the *M. tuberculosis* isolates have identical Spoligotyping or MIRU-VNTR patterns. The dendrograms based on the Spoligotyping and 24 VNTR loci data were conducted using UPGMA protocol. The Hunter-Gaston discrimination index (HGDI) for the 24 VNTR loci was calculated using the previously reported formula [18]. The HGDI of each VNTR locus was calculated by using the web based VNTR Diversity and Confidence Extractor (V-DICE <http://www.hpa-bioinformatics.org.uk/cgi-bin/DICI/DICI.pl>).

2.6. Statistical Analysis. Chi-squared tests were used to compare the proportions of DR-TB and the distribution of genotypes in different groups. All tests of significance were two-sided and the significant threshold was set at 0.05.

3. Results

3.1. Demographic Characters. All of the 265 clinical isolates were cultured from sputum samples of different TB patients in Zigong CDC, Sichuan province, in 2010. And for the 265 TB patients, 44 (16.60%) were previously treated, 139 (52.45%) were in the 30- and 60-year-old age group, 39 (14.72%) female and 209 (78.87%) male patients with other 17 (6.41%) gender uncertain patients, and no difference about these factors between Beijing and non-Beijing family strains was found at $p < 0.05$ level (Table 1).

3.2. Spoligotyping. Based on the Spoligotyping results, all of the 265 isolates were subdivided into 64 Spoligotypes, among which 49 patterns were unique whereas other 216 isolates were grouped into 15 clusters containing 2 to 128 isolates. Following comparison with the SITVITWEB database, 239 isolates were clustered into 40 shared international types (SITs), whereas the other 26 isolates that had not been described in the database were referred to as "Orphan." Family assignment revealed that 143 strains had the classical Beijing family-specific or Beijing family-like Spoligotyping pattern. We treated all these 143 isolates as Beijing family strains (Figure 1). In total, the Beijing family is clearly the most prevalent genetic family in our research setting (143/265, 53.96%), followed by T family (80/265, 30.19%) and H family (9/265, 3.40%).

3.3. MIRU-VNTR Typing. Exploitable VNTR results (defined as less than 4 missing values) were obtained for 262 isolates out of 265 isolates. Using the 24 polymorphic loci, all of the 262 *M. tuberculosis* isolates were identified to 257 different VNTR types and 3 main clusters. A total of 253 isolates had unique profiles. The remaining 9 isolates formed 4 clusters (2 to 3 isolates) (Figure 2). The total Hunter-Gaston

TABLE 1: Demographic data of the 265 isolates in Zigong, China.

| Factors | Whole (<i>n</i> = 265) | Beijing (<i>n</i> = 143) | Non-Beijing (<i>n</i> = 122) | χ^2 * (<i>p</i> value) |
|-----------------|----------------------------|------------------------------|----------------------------------|------------------------------|
| Sex | | | | |
| Female | 39 | 25 | 14 | 1.6036 (0.2054) |
| Male | 209 | 111 | 98 | |
| Unknown | 17 | 7 | 10 | |
| Age | | | | |
| <30 | 27 | 19 | 8 | 6.9592 (0.0732) |
| ≥30 | 139 | 79 | 60 | |
| ≥60 | 69 | 30 | 39 | |
| ≥75 | 9 | 6 | 3 | |
| Unknown | 21 | 9 | 12 | |
| Treated before? | | | | |
| Yes | 44 | 25 | 19 | 0.1732 (0.6773) |
| No | 221 | 118 | 103 | |

*: evaluated the difference between Beijing and non-Beijing family strains.

TABLE 2: HGDI of the 24 MIRU-VNTR loci for the whole sample, Beijing and non-Beijing family isolates.

| Locus | Whole (<i>n</i> = 265) | Beijing (<i>n</i> = 143) | Non-Beijing (<i>n</i> = 122) |
|---------|----------------------------|------------------------------|----------------------------------|
| ETRA | 0.594 | 0.262 | 0.606 |
| ETRB | 0.374 | 0.068 | 0.504 |
| ETRC | 0.110 | 0.069 | 0.159 |
| ETRD | 0.303 | 0.028 | 0.540 |
| ETRE | 0.592 | 0.245 | 0.202 |
| MIRU02 | 0.030 | 0.000 | 0.065 |
| MIRU10 | 0.576 | 0.263 | 0.242 |
| MIRU16 | 0.283 | 0.181 | 0.392 |
| MIRU20 | 0.246 | 0.240 | 0.254 |
| MIRU23 | 0.124 | 0.108 | 0.143 |
| MIRU24 | 0.015 | 0.014 | 0.017 |
| MIRU26 | 0.739 | 0.570 | 0.598 |
| MIRU27 | 0.328 | 0.028 | 0.523 |
| MIRU39 | 0.521 | 0.082 | 0.097 |
| MIRU40 | 0.551 | 0.485 | 0.619 |
| Mtub04 | 0.620 | 0.363 | 0.622 |
| Mtub21 | 0.806 | 0.710 | 0.571 |
| Mtub29 | 0.096 | 0.069 | 0.127 |
| Mtub30 | 0.496 | 0.119 | 0.182 |
| Mtub34 | 0.095 | 0.042 | 0.157 |
| Mtub39 | 0.511 | 0.267 | 0.647 |
| QUB11b | 0.854 | 0.745 | 0.796 |
| QUB26 | 0.723 | 0.571 | 0.821 |
| QUB4156 | 0.052 | 0.081 | 0.017 |

discrimination index (HGDI) for all of the 263 isolates was 0.9995, and that of Beijing and non-Beijing family isolates was 0.9987 and 0.9992, respectively; the HGDI of each of the 24 VNTR loci were shown in Table 2.

3.4. Drug-Resistant Phenotypic Profiles. Among all the 265 isolates, 72 (27.17%) were resistant to at least one of the eight tested anti-TB drugs, and other 193 (72.83%) were sensitive. Of the 72 drug-resistant isolates, 48 (18.11%) were INH-resistant, 40 (15.09%) were RIF-resistant, 32 (12.08%) were SM-resistant, 17 (6.42%) were EMB-resistant, 32 (12.08%) were OFLX-resistant, and 36 (13.58%) of those drug-resistant strains were MDR-TB; no XDR-TB was found in this study (Table 3). When grouped by gender or age group like previous report [8], no relationship was found with the isolate’s character of drug resistance. But having been treated before may play an important role with the drug resistance, not only for the drug-resistant TB ($\chi^2 = 11.2680$, $p < 0.05$) but also for the MDR-TB ($\chi^2 = 18.8983$, $p < 0.05$) (Table 4).

3.5. Relationship between Beijing Family and Drug Resistance. According to the demographic data, no matter whether they are grouped by gender, age groups, or even treatment history, there was no difference of the infection rate between Beijing and non-Beijing family isolates observed (Table 1). And for the Spoligotyping patterns and DST results, of the 143 Beijing family isolates, 41 (28.67%) were drug-resistant to at least one of the eight anti-TB drugs of which 23 (16.08%) were MDR-TB. While, in the 122 non-Beijing isolates, 31 (25.41%) were drug-resistant and 13 (10.66%) were MDR-TB, no difference of the drug resistance proportion was found between Beijing and non-Beijing family isolates (Table 3).

4. Discussion

Previous research report has reviewed that Beijing family is prevalent (57.89%) in Sichuan province [6], and according to the Spoligotyping results in this study, the proportion of Beijing family is about 54.34% in southern Sichuan, a little lower than that of the previous report. This may be because our study was just focused on the prevalence isolates in Zigong, and more isolates were included in this study which

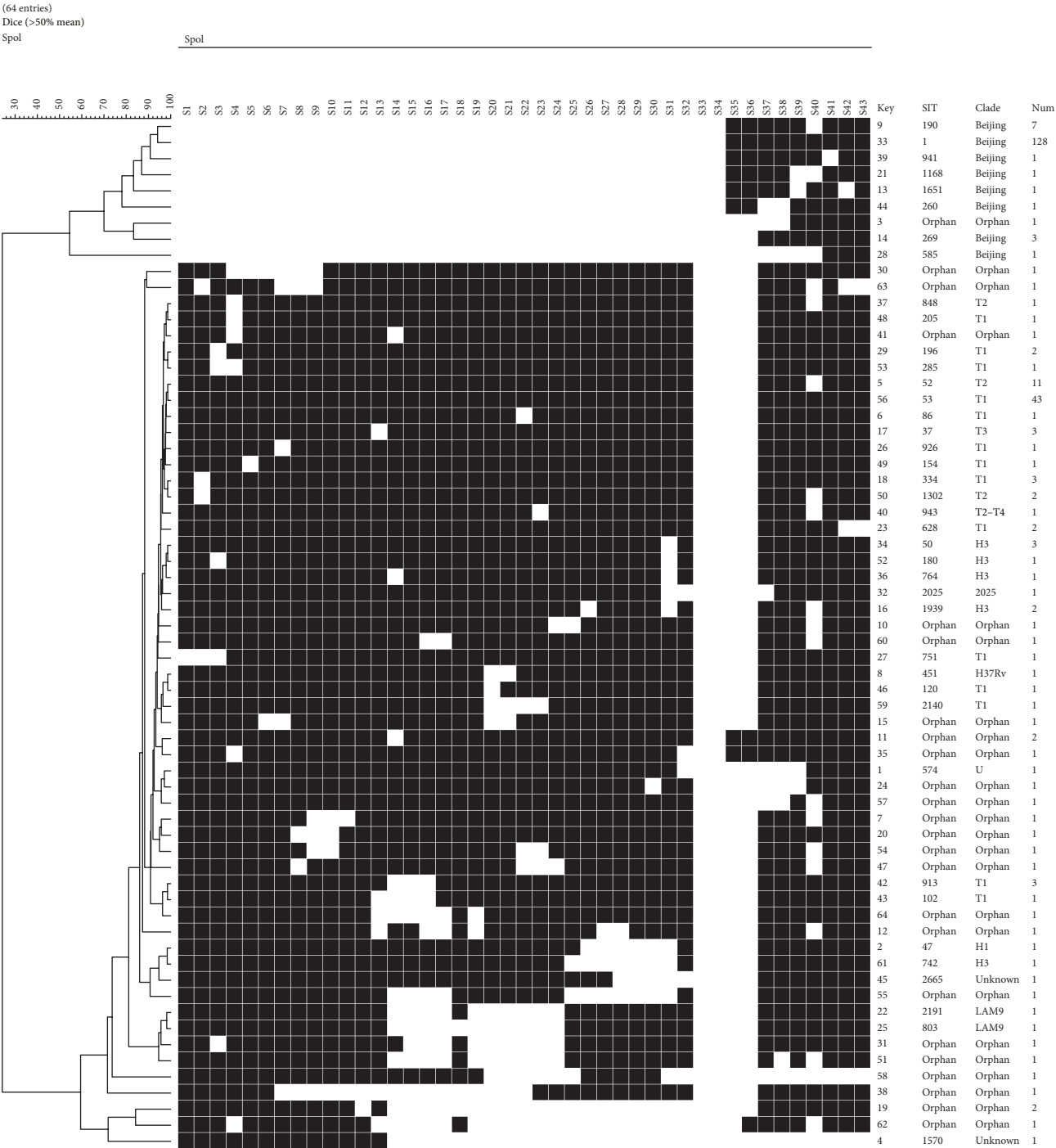


FIGURE 1

could afford much more representativeness. But this still conforms to the former reports that the most prevalent isolate here was Beijing family strains and the highest prevalence of Beijing family was found in northern China, followed by central and southern China [6, 7]. Besides, T genotype was the second major epidemic genotype circulating in this area and was also consistent with the previous study [19]. This may indicate that non-Beijing family strains also played an important role in the high tuberculosis burden in Sichuan

and some other provinces that are located in southern China [6, 7].

Spoligotyping has been considered as the gold standard for Beijing family identification, as it is simple and efficient. But the discrimination of Spoligotypes for Beijing family is very low and it cannot further differentiate Beijing family strains, as Beijing family strains always show characteristic homological Spoligotyping pattern [20]. For the 24 loci MIRU-VNTR typing method, even the discrimination of

TABLE 3: Prevalence of anti-TB drug resistance among 265 isolates in Sichuan province*.

| Anti-TB drugs | Whole (n = 265) | Beijing (n = 143) | Non-Beijing (n = 122) | χ^2 (p value) |
|--------------------------------|--------------------|----------------------|--------------------------|--------------------|
| Isoniazid | 18.11% (n = 48) | 19.44% (n = 28) | 16.53% (n = 20) | 0.45 (0.5019) |
| Rifampicin | 15.09% (n = 40) | 18.75% (n = 27) | 10.74% (n = 13) | 3.48 (0.0623) |
| Streptomycin | 12.08% (n = 32) | 13.89% (n = 20) | 9.92% (n = 12) | 1.07 (0.3014) |
| Ethambutol | 6.42% (n = 17) | 7.64% (n = 11) | 4.96% (n = 6) | 0.84 (0.3583) |
| Ofloxacin | 12.08% (n = 32) | 9.03% (n = 13) | 15.70% (n = 19) | 2.61 (0.1065) |
| Resistant to any anti-TB drugs | 27.17% (n = 72) | 28.47% (n = 41) | 25.62 % (n = 31) | 0.35 (0.5519) |
| MDR-TB | 13.58% (n = 36) | 15.97% (n = 23) | 10.74% (n = 13) | 1.65 (0.1986) |

*: all 265 isolates were sensitive to other three kinds of anti-TB drugs, and Pearson χ^2 and p value were evaluated between the Beijing and non-Beijing family strains.

TABLE 4: Prevalence of anti-TB drug resistance among 265 isolates by factors.

| Factors | Sensitive TB (n = 193) | Any drug resistance TB (n = 72) | χ^2 (p value) | Non-MDR-TB (n = 229) | MDR-TB (n = 36) | χ^2 (p value) |
|----------|---------------------------|---------------------------------------|--------------------|-------------------------|--------------------|--------------------|
| Sex | | | | | | |
| Female | 29 | 10 | 0.1097 (0.7405) | 34 | 5 | 0.0095 (0.9225) |
| Male | 150 | 59 | | 181 | 28 | |
| Unknown | 14 | 3 | | 14 | 3 | |
| Age | | | | | | |
| <30 | 22 | 5 | 4.1444* (0.2402) | 23 | 4 | 4.8455* (0.1579) |
| ≥30 | 93 | 46 | | 115 | 24 | |
| ≥60 | 52 | 17 | | 64 | 5 | |
| ≥75 | 8 | 1 | | 9 | 0 | |
| Unknown | 18 | 3 | | 18 | 3 | |
| Treated? | | | | | | |
| Yes | 23 | 21 | 11.2680 (0.0008) | 29 | 15 | 18.8983 (0.0000) |
| No | 170 | 51 | | 200 | 21 | |

*: evaluated by the Fisher's exact test.

each locus is different, but when we combined them together a high resolution was acquired (Table 2). As the MIRU-VNTR genotyping is considered to be faster and easier to perform, the results are in digital format which can be easily compared between laboratories. Thus, if conditions allow, we suggest Spoligotyping and MIRU-VNTR genotyping should be taken out at the same time, to investigate in more detail characters of the selected isolates. But for some loci the HGDI were so low; different countries could select specific loci for genotyping. And for the three main clusters of the UPGMA dendrogram in Figure 2, one formed by Beijing family isolates and other two within non-Beijing family isolates, but no difference was found between them for difference factors, like drug resistance and patient age group.

In our study, 27.17% (n = 72) of the isolates were resistant to at least one of the eight anti-TB drugs, and 13.58%

(n = 36) were MDR-TB; no XDR-TB was found in this study. These have some difference with the whole nation's status [4]. Firstly, we focused on the target area here in Sichuan province; the demographic and *M. tuberculosis* infectious status in this area may have some differences with those of the whole nation. Secondly, the isolates selected into our study may also have some differences with that of the previously report. Also in our study we found that the treatment history may play an important role with the drug resistance, not only for drug-resistant TB but also for MDR-TB (Table 4); thus the management of TB patients should be strengthened. And as the drug-resistant TB may be the most important challenge to achieve the goals of stopping TB, much more attention should be paid to it.

As Beijing family strains were prevalent in this area (143/265, 53.96%) and in some reports, Beijing family has

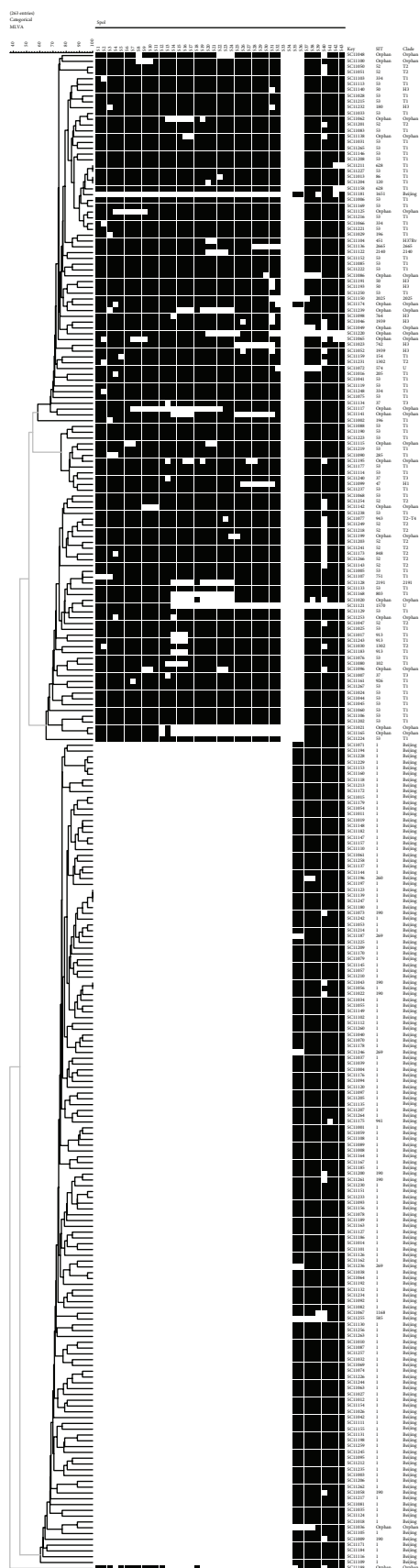


FIGURE 2

been considered to be related to TB outbreak and drug resistance [21–24], while some are not [25–27]. Referring to our study, the chi-square analysis showed that there was no significant difference between the Beijing and non-Beijing family strains in terms of their drug resistance proportion and patient age group. This may be because Beijing family is the genetic aspect but the drug resistance is a biological aspect of one *M. tuberculosis* isolate, and there is no constant relationship between them.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors thank the staffs of Zigong Center for Disease Control and Prevention for their excellent contribution to this study. This work was funded by the Chinese National Key Program of Mega Infectious Disease Project 2013ZX10003006-002-001.

References

- [1] World Health Organization, *Global Tuberculosis Report 2013*, World Health Organization, Geneva, Switzerland, 2013.
- [2] N. R. Gandhi, P. Nunn, K. Dheda et al., “Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis,” *The Lancet*, vol. 375, no. 9728, pp. 1830–1843, 2010.
- [3] World Health Organization, *Multidrug and Extensively Drug-Resistant TB (M/XDR-TB): 2010 Global Report on Surveillance and Response*, World Health Organization, 2010.
- [4] Y. Zhao, S. Xu, L. Wang et al., “National survey of drug-resistant tuberculosis in China,” *The New England Journal of Medicine*, vol. 366, no. 23, pp. 2161–2170, 2012.
- [5] Y. Jiang, H.-C. Liu, H. Zheng et al., “19-VNTR loci used in genotyping Chinese clinical *Mycobacterium tuberculosis* complex strains and in association with spoligotyping,” *Journal of Basic Microbiology*, vol. 53, no. 7, pp. 562–580, 2013.
- [6] H. Dong, Z. Liu, B. Lv et al., “Spoligotypes of *Mycobacterium tuberculosis* from different provinces of China,” *Journal of Clinical Microbiology*, vol. 48, no. 11, pp. 4102–4106, 2010.
- [7] H. Dong, L. Shi, X. Zhao et al., “Genetic diversity of *Mycobacterium tuberculosis* isolates from Tibetans in Tibet, China,” *PLoS ONE*, vol. 7, no. 3, Article ID e33904, 2012.
- [8] Q. Liu, D. Yang, W. Xu et al., “Molecular typing of *Mycobacterium tuberculosis* isolates circulating in Jiangsu province, China,” *BMC Infectious Diseases*, vol. 11, article 288, 2011.
- [9] B. Lu, P. Zhao, B. Liu et al., “Genetic diversity of *Mycobacterium tuberculosis* isolates from Beijing, China assessed by Spoligotyping, LSPs and VNTR profiles,” *BMC Infectious Diseases*, vol. 12, article 372, 2012.
- [10] Q. Yu, Y. Su, B. Lu et al., “Genetic diversity of *Mycobacterium tuberculosis* isolates from Inner Mongolia, China,” *PLoS ONE*, vol. 8, no. 5, Article ID e57660, 2013.
- [11] L. Wang, S. Cheng, M. Chen et al., “The fifth national tuberculosis epidemiological survey in 2010,” *Chinese Journal of Antituberculosis*, vol. 34, pp. 485–508, 2012.

- [12] J.-H. Guo, W.-L. Xiang, Q.-R. Zhao et al., "Molecular characterization of drug-resistant *Mycobacterium tuberculosis* isolates from Sichuan Province in China," *Japanese Journal of Infectious Diseases*, vol. 61, no. 4, pp. 264–268, 2008.
- [13] J. Kamerbeek, L. Schouls, A. Kolk et al., "Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology," *Journal of Clinical Microbiology*, vol. 35, no. 4, pp. 907–914, 1997.
- [14] C. Demay, B. Liens, T. Burguière et al., "SITVITWEB—a publicly available international multimer database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology," *Infection, Genetics and Evolution*, vol. 12, no. 4, pp. 755–766, 2012.
- [15] P. Supply, E. Mazars, S. Lesjean, V. Vincent, B. Gicquel, and C. Locht, "Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome," *Molecular Microbiology*, vol. 36, no. 3, pp. 762–771, 2000.
- [16] World Health Organization, *Guidelines for Surveillance of Drug Resistance in Tuberculosis*, World Health Organization, 2009.
- [17] World Health Organization, *Guidelines for the Programmatic Management of Drug-Resistant Tuberculosis-2011 Update*, World Health Organization, Geneva, Switzerland, 2011.
- [18] P. R. Hunter and M. A. Gaston, "Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity," *Journal of Clinical Microbiology*, vol. 26, no. 11, pp. 2465–2466, 1988.
- [19] K. Wan, J. Liu, Y. Hauck et al., "Investigation on *Mycobacterium tuberculosis* diversity in china and the origin of the Beijing clade," *PLoS ONE*, vol. 6, no. 12, Article ID e29190, 2011.
- [20] D. van Soolingen, L. Qian, P. E. W. de Haas et al., "Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia," *Journal of Clinical Microbiology*, vol. 33, no. 12, pp. 3234–3238, 1995.
- [21] K. Kremer, J. R. Glynn, T. Lillebaek et al., "Definition of the Beijing/W lineage of *Mycobacterium tuberculosis* on the basis of genetic markers," *Journal of Clinical Microbiology*, vol. 42, no. 9, pp. 4040–4049, 2004.
- [22] Y. Hu, X. Ma, E. A. Graviss, W. Wang, W. Jiang, and B. Xu, "A major subgroup of Beijing family *Mycobacterium tuberculosis* is associated with multidrug resistance and increased transmissibility," *Epidemiology and Infection*, vol. 139, no. 1, pp. 130–138, 2011.
- [23] S. Purwar, S. Chaudhari, V. M. Katoch et al., "Determination of drug susceptibility patterns and genotypes of *Mycobacterium tuberculosis* isolates from Kanpur district, North India," *Infection, Genetics and Evolution*, vol. 11, no. 2, pp. 469–475, 2011.
- [24] T. Brown, V. Nikolayevskyy, P. Velji, and F. Drobniewski, "Associations between *Mycobacterium tuberculosis* strains and phenotypes," *Emerging Infectious Diseases*, vol. 16, no. 2, pp. 272–280, 2010.
- [25] A. Ani, T. Bruvik, Y. Okoh et al., "Genetic diversity of *Mycobacterium tuberculosis* complex in Jos, Nigeria," *BMC Infectious Diseases*, vol. 10, article 189, 2010.
- [26] M. K. Kai, W. Y. Chi, W. T. Lai et al., "Utility of mycobacterial interspersed repetitive unit typing for differentiating multidrug-resistant *Mycobacterium tuberculosis* isolates of the Beijing family," *Journal of Clinical Microbiology*, vol. 43, no. 1, pp. 306–313, 2005.
- [27] J. Zhang, L. Mi, Y. Wang et al., "Genotypes and drug susceptibility of *Mycobacterium tuberculosis* Isolates in Shihezi, Xinjiang Province, China," *BMC Research Notes*, vol. 5, article 309, 2012.