New groups of CRISPR-associated protein 1 (Cas1) from hot springs

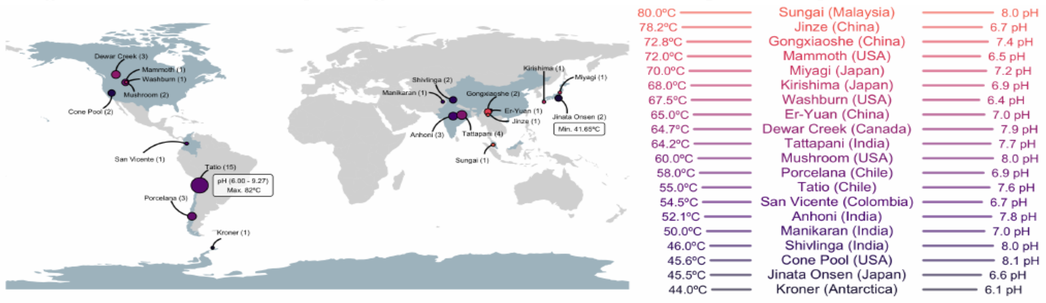
Oscar Salgado[1] & Beatriz Díez[1,2,3].
[1]Bioinformatics Laboratoy, Education Faculty, Adventist University of Chile, Chile.
[2]Departamento de Genética Molecular y Microbiología, Pontificia Universidad Católica de Chile, Chile.
[3]Center for Climate and Resilience Research (CR)2, Chile.
bdiez@bio.puc.cl

INTRODUCTION

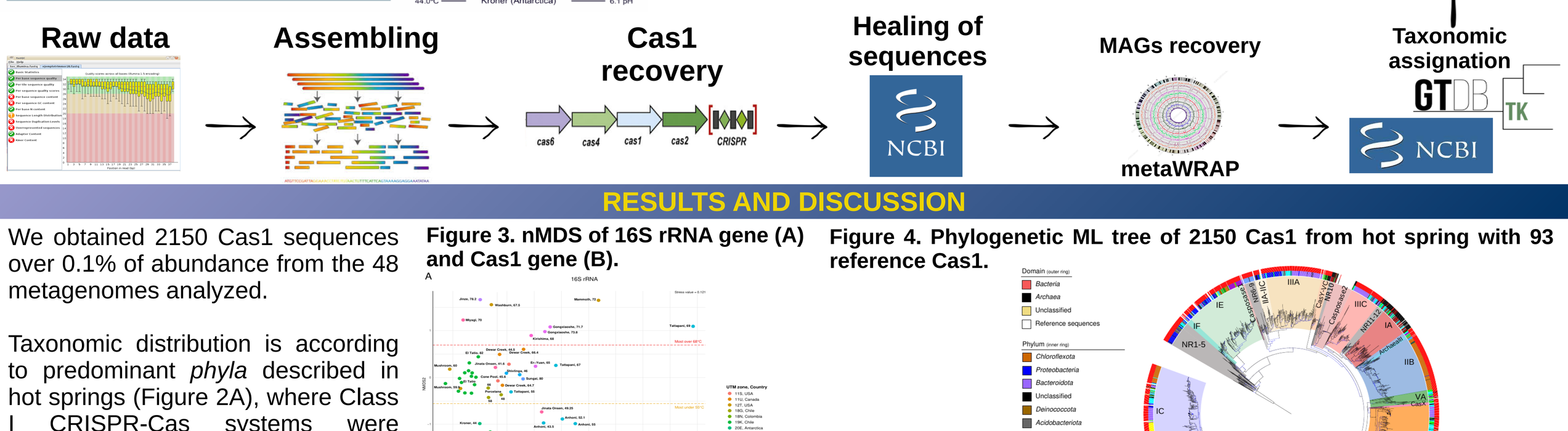
Microorganisms threaten viral infections, which have caused the maintenance of molecular immunity mechanisms. The CRISPR-Cas systems (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated proteins, Cas) perform the recognition and degradation of invading foreign nucleic acids as viruses. There are different types of CRISPR-Cas systems, and each one has a set of different Cas proteins. These proteins are highly diverse and mobile. Nevertheless, the Cas1 protein is recognized as the most conserved Cas and allows for to description of the represented groups of the CRISPR-Cas systems. On the other hand, some Cas1 sequences do not belong to CRISPR-Cas systems; instead, they are part of special transposons, called casposons due to Cas1 presence, an enzyme called casposase. CRISPR-Cas systems are more prevalent in microorganisms living in high-temperature environments; hence these environments are a perfect model for studying the diversity of CRISPR-Cas systems through Cas1 analyses due to the low complexity of hot springs communities compared with mesophilic environments. We used a phylogenetic and similarity approach to describe the ecological diversity of Cas1 proteins from hot springs. Our results show the novelty of Cas1 sequences from hot springs of the world.

MATERIALS AND METHODS

Figure 1. 48 hot springs used in this study with pH and



80.0°C	Sungai (Malaysia)	8.0 pH
78.2°C	Jinze (China)	6.7 pH
72.8°C	Gongxiaoshe (China)	7.4 pH
72.0°C	Mammoth (USA)	6.5 pH
70.0°C	Miyagi (Japan)	7.2 pH
68.0°C	Kirishima (Japan)	6.9 pH
67.5°C	Washburn (USA)	6.4 pH
65.0°C	Er-Yuan (China)	7.0 pH
64.7°C	Dewar Creek (Canada)	7.9 pH
64.2°C	Tattapani (India)	7.7 pH
60.0°C	Mushroom (USA)	8.0 pH
58.0°C	Porcelana (Chile)	6.9 pH
55.0°C	Tatio (Chile)	7.6 pH
54.5°C	San Vicente (Colombia)	6.7 pH
52.1°C	Anhoni (India)	7.8 pH
50.0°C	Manikaran (India)	7.0 pH
46.0°C	Shivlinga (India)	8.0 pH
45.6°C	Cone Pool (USA)	8.1 pH
45.5°C	Jinata Onsen (Japan)	6.6 pH
44.0°C	Kroner (Antarctica)	6.1 pH

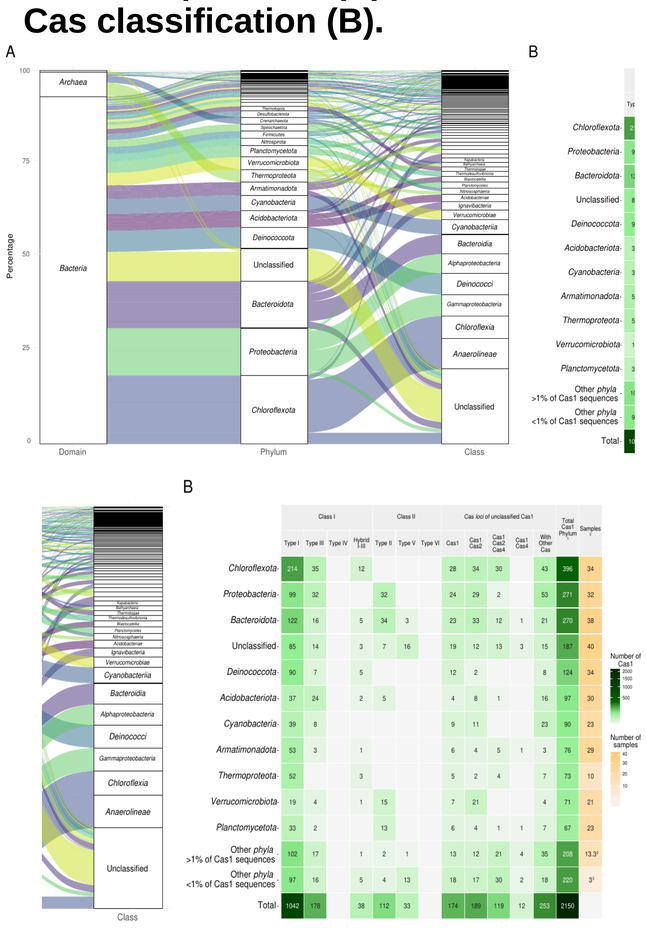


RESULTS AND DISCUSSION

We obtained 2150 Cas1 sequences over 0.1% of abundance from the 48 metagenomes analyzed.

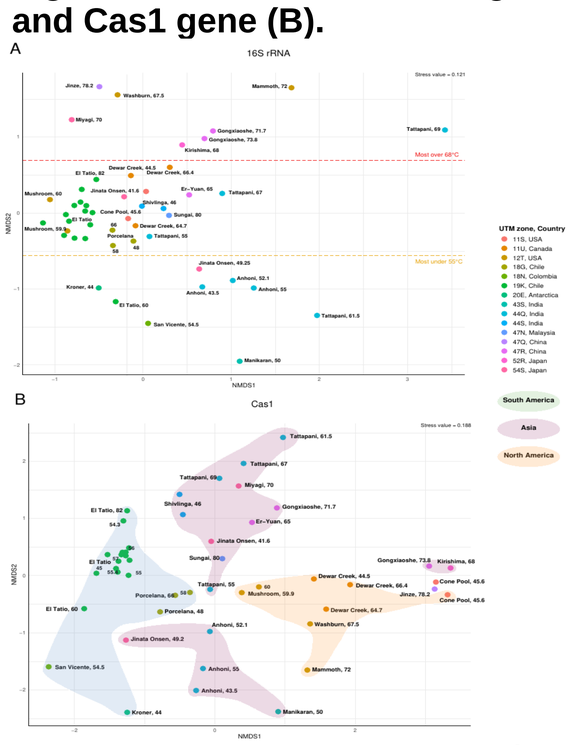
Taxonomic distribution is according to predominant *phyla* described in hot springs (Figure 2A), where Class I CRISPR-Cas systems were observed (Figure 2B).

Figure 2. Taxonomic distribution of Cas1 sequences (A) and CRISPR-Cas classification (B).



The Bray-Curtis diversity index of the 16S rRNA reveals that predominant taxa of hot springs are less dissimilar than Cas1 sequences (Figure 3), which suggests endemic traits of this gene and indicates a specific virus-host relationship in hot springs.

Figure 3. nMDS of 16S rRNA gene (A) and Cas1 gene (B).

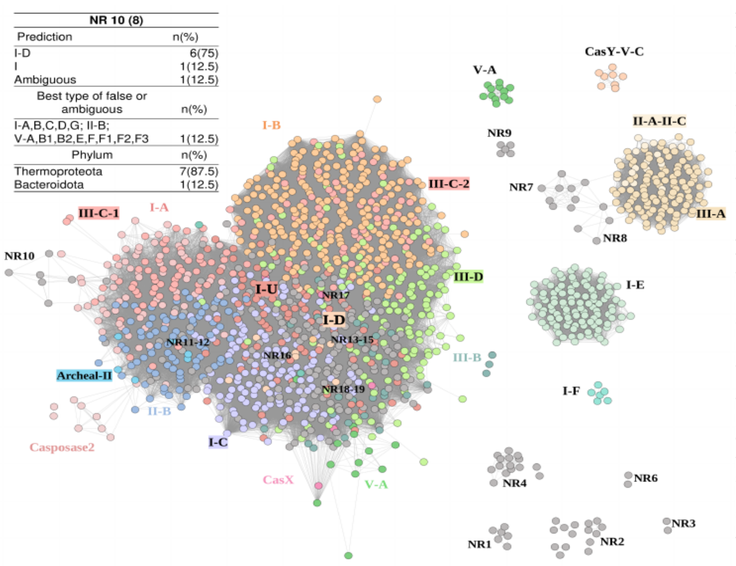


The phylogenetic tree reveals that Cas1 tends to group according to CRISPR-Cas systems type. However, several clades do not group with reference Cas1 sequences from databases (NR, Figure 4). These clades are composed of sequences of several taxa and tend to be at the root of significant clades, which is also observed for Cas1 of recently discovered CRISPR-Cas systems or casposons Cas1.

Th similarity of Cas1 through network analyses shows that most Class I Cas1 shapes a big module separated from Class II sequences and NR sequences (Figure 5). This suggests that Cas1 proteins from hot springs are new for databases and could represent new CRISPR-Cas systems.

Figure 4. Phylogenetic ML tree of 2150 Cas1 from hot spring with 93 reference Cas1.

Figure 5. Network analyses of Cas1 from hot spring with 93 reference Cas1.



NR 11-19 (252)	
Prediction	n(%)
False	143(56.7)
III-A	50(19.8)
I	30(12)
III-B	13(5.1)
III-D	6(2.3)
I-A	6(2.3)
Ambiguous	3(1.2)
I-B	1(0.4)
Best type of false or ambiguous	n (%)
I-A	136(54)
III-B	4(1.5)
I-A,B,C,D,G; II-B; V-A,B1,B2,E,F,F1,F2,F3	3(1.2)
III-A	3(1.2)
Phylum	n(%)
Chloroflexota	124(49.2)
Unknown	26(10.3)
Desulfobacterota	10(4)
Crenarchaeota	9(3.5)
Spirochaetota	9(3.5)
Cyanobacteria	8(3.1)
Deferribacterota	7(2.7)
Myxococcota	7(2.7)
Thermoproteota	6(2.3)
Acidobacteriota	4(1.5)
Bipolaricaulota	4(1.5)
Calditrichota	4(1.5)
Planctomycetota	4(1.5)
Armatimonadota	3(1.2)
Firmicutes	3(1.2)
Halobacteriota	3(1.2)
Nitrospirota	3(1.2)
Proteobacteria	3(1.2)
Thermotogota	3(1.2)
Phyla under 1%	12(4.7)

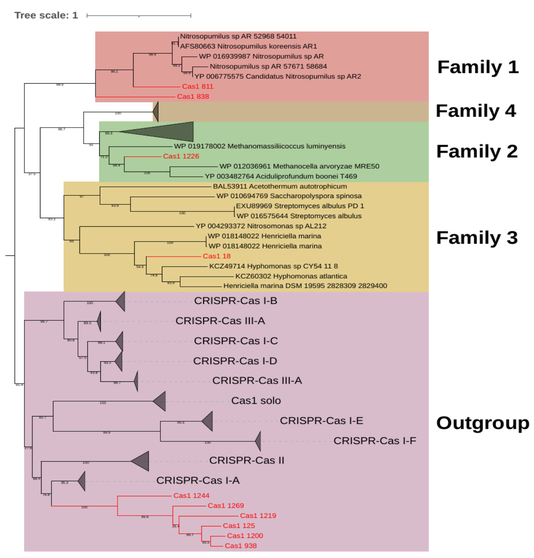
NR 10 (8)	
Prediction	n(%)
I-D	6(75)
I	1(12.5)
Ambiguous	1(12.5)
Best type of false or ambiguous	n(%)
I-A,B,C,D,G; II-B; V-A,B1,B2,E,F,F1,F2,F3	1(12.5)
Phylum	n(%)
Thermoproteota	7(87.5)
Bacteroidota	1(12.5)

NR 7-9 (26)	
Prediction	n(%)
False	15(57.7)
II	11(42.3)
Best type of false or ambiguous	n (%)
II-A	13(50)
I-A	2(7.7)
Phylum	n (%)
Verrucomicrobiota	25(96.1)
Unknown	1(3.8)

NR 1-6 (72)	
Prediction	n(%)
False	45(62.5)
III-A	15(20.8)
I	4(5.5)
Ambiguous	3(4.1)
III-B	3(4.1)
Hybrid(III-A,III-B)	1(1.3)
I-A	1(1.3)
Best type of false or ambiguous	n (%)
I-A	36(50)
III-B	9(12.5)
I-A,B,D,E,F,G; III-A,B,C,D	3(4.1)
Phylum	n (%)
Proteobacteria	36(50)
Acidobacteriota	16(22.2)
Nitrospirota	9(12.5)
Unknown	7(9.7)
Verrucomicrobiota	3(4.1)
UBA2233	1(1.3)

In-depth phylogenetic and syntenic analyses of Cas1 that group close to casposase (Cas1-solo) indicate that hot spring harbor new casposons, which are located in a branch distant from described casposase families and vestigial Cas1-solo (Figure 6).

Figure 6. Phylogenetic ML tree of Cas1-solo from hotsprings.



CONCLUSIONS

- Cas1 genes from hot springs are represented in predominant *phyla* of these environments.
- Cas1 from hot springs tends to be specific to each geographical location.
- Novel phylogenetic clades are observed using Cas1 from hot springs and reference sequences, which suggest new CRISPR-Cas systems.
- Casposases from hot springs reveal a new family of casposons.