

Genomics and Molecular Phylogenetics Tree Analysis of *Actinopolyspora Iraquiensis*

Talal Sabhan Salih

talal.salih@uomosul.edu.iq

Department of Biophysics,
College of Science, University of Mosul, Iraq

Raghad Riyadh Shafeek

raghad.riyadh@uomosul.edu.iq

College of Science,
University of Mosul, Iraq

Abstract

Actinopolyspora iraqiensis IQ-H1 is a novel strain of actinobacteria isolated from extremely halophilic soil samples in Iraq. The whole-genome sequence of this strain is deposited in the National Center for Biotechnology Information (NCBI) GenBank under the accession number NZ_AICW01000000. In this study, the genome features and the molecular phylogenetic tree of *Act. iraqiensis* IQ-H1 are analyzed. The RAST tool was used for genome annotation. The genomic features were elucidated using QUAST tool. The circular genome map, and the core and pan-genome map of *Act. iraqiensis* IQ-H1 was generated using CGView and the GView tools respectively. The JSpeciesWS server was used for the tetranucleotide signature analysis and the REALPHY server was utilized for the construction of the whole genome sequence based phylogenetic tree. The genome size of the strain was around 4.0 Mpb and the number of contigs was 110 with a GC content of 70.46%. The core genome of *Act. iraqiensis* IQ-H1 was estimated to be 2.2 Mpb. Based on z-scores of the tetranucleotide signature analysis, *Act. halophila* DSM 43834, *Act. mortivallis* DSM 44261 and *Act. saharensis* DSM 45459 were the most relative strains to *Act. iraqiensis* IQ-H1 with z-scores 0.99784, 0.98943 and 0.99789 respectively. Based on the phylogenetic tree constructed from the whole genome sequences, *Act. iraqiensis* IQ-H1 was the most closely related to *Act. saharensis* DSM 45459, *Act. halophila* DSM 43834 and *Act. mortivallis* DSM 44261. The results suggest that the web-based bioinformatics tools such as QUAST, CGView, GView, JSpeciesWS and REALPHY can be utilized for the analysis of the genomic features of *Act. iraqiensis* IQ-H1 and other species of the genus *Actinopolyspora*.

Keyword: *Actinopolyspora iraqiensis*, Phylogenetics Tree, Genomic Features, Tetranucleotide Signature.

1. Introduction

The genus of *Actinopolyspora* was proposed for the first time in 1975 by Gochnauer and his colleagues [1]. The genus currently encompasses 13 species with validly published names *Act. halophila*, *Act. mortivallis*, *Act. iraqiensis*, *Act. alba*, *Act. erythraea*, *Act. xinjiangensis*, *Act. algeriensis*, *Act. lacussalsi*, *Act. mzabensis*, *Act. saharensis*, *Act. righensis*, *Act. biskrensis* and *Act. salinaria* [2, 3]. Species belong to the genus *Actinopolyspora* are extremely halophilic which can grow in saturated NaCl up to 20% [4]. The strains of this genus are Gram positive bacteria that belong to the phylum Actinobacteria with high GC content (67-70%) of the DNA [2]. Like other members of Actinobacteria which are well-

For more information about the Conference please visit the websites:

<http://ihicps.com/>

Biology | 38

known for producing bioactive compounds [5, 6], some *Actinopolyspora* has been reported to produce antibacterial bioactive compounds [7, 8].

In our study, the whole genome sequence of *Act. iraqiensis* strain IQ-H1, which isolated from extremely saline soil samples in Iraq [9], along with eight *Actinopolyspora* strains whose whole genome sequences are available in the National Center for Biotechnology Information (NCBI) were utilized. The genomics and molecular phylogenetics analyses utilizing genomic comparative programs and tools showed that *Act. iraqiensis* IQ-H1 has unique genomic features and has different genomic characteristics at the strain level than other related *Actinopolyspora* species.

2. Materials Method

2.1. Whole genome sequences of *Actinopolyspora*

As the main objective of this research is to genomics analysis of *Act. iraqiensis* IQ-H1. The genomic reference sequences of *Actinopolyspora* genus as well as the whole genome sequence of *Act. iraqiensis* were utilised. The genome sequences of *Act. halophila* DSM 43834, *Act. mortivallis* DSM 4426, *Act. erythraea* YIM 90600, *Act. righensis* DSM 4550, *Act. alba* DSM 45004, *Act. saharensis* DSM 45459, *Act. mzabensis* DSM 45460 and *Act. xinjiangensis* DSM 46732 were obtained from the National Center for Biotechnology Information (NCBI) Genbank database as of March 2020. Genome sequences along with their accession numbers, genome sizes, number of contigs, and GC contents are listed in **Table 1**. The whole-genome sequences were downloaded and stored in a fasta format for further analyses.

Table 1. *Actinopolyspora* reference whole genome sequences used in this study.

Genome	Accession Number	Size (Mbp)	Number of Contigs	GC(%)
<i>Act. halophila</i> DSM 43834	AQUI000000000	5.25	1	68.0
<i>Act. mortivallis</i> DSM 44261	NZ_AQZN000000000	4.23	18	68.8
<i>Act. erythraea</i> YIM 90600	NZ_CP022752	5.24	1	68.8
<i>Act. righensis</i> DSM 45501	NZ_FPAT000000000	4.92	23	67.5
<i>Act. alba</i> DSM 45004	NZ_FOMZ000000000	5.23	42	67.6
<i>Act. saharensis</i> DSM 45459	NZ_FNKO000000000	4.68	2	69.5
<i>Act. mzabensis</i> DSM 45460	NZ_FNFM000000000	5.00	25	67.7
<i>Act. xinjiangensis</i> DSM 46732	NZ_FNJR000000000	5.03	34	68.4

2.2. Genome features of *Act. iraqiensis* IQ-H1

To study the genome features of *Act. iraqiensis* IQ-H1, the whole genome sequence of this strain was first uploaded to the Rapid Annotation using Subsystem Technology tool (RAST) [10] for annotation. The annotated genome was then sent to QUASt tool [11] to elucidate the unique features. rRNA annotation was done using tRNAscan -SE v2 .0 program [12]. The circular genome map of *Act. iraqiensis* IQ-H1 was performed using CGView Comparison Tool [13] as the annotated files produced by RAST were in the GeneBank (.gbk) and Gene-Finding (.gff) formats *utilized*.

2.3. Core and Pan-Genome Comparative Analysis

For more information about the Conference please visit the websites:

<http://ihicps.com/>

For the generation of the core and pan-genome map, the GView tool [14] was used. The *Act. iraqiensis* IQ-H1 genome along with all the reference genome sequences in GenBank format were uploaded to the server. The *Act. erythraea* YIM 90600 was selected as a seed genome and the other genomes were compared to the seed to locate the unique regions. The seed is incrementally built up with the unique features of the queries to become the pan-genome. A BLAST atlas was created to display the presence or absence of features within the query genomes compared to the pan-genome.

2.4. Tetranucleotide Signature Analysis

The tetranucleotide signature analysis computes correlation coefficients between tetranucleotide usage patterns of DNA sequences, which can be used as an indicator of bacterial genome sequences relatedness. The calculation of tetranucleotide frequencies for each genome sequence was performed according to [15] through the JSpeciesWS server [16]. In brief, a fragment of DNA sequence with 4 bases can be transformed to an array of 256 possible tetranucleotide patterns and their corresponding expected frequencies are computed. The differences between frequencies and expected values are transformed into Z-scores for each tetranucleotide.

2.5. Whole Genome Sequence Based Phylogenetic Tree

For construction of a maximum likelihood phylogenetic tree based on whole genome sequences, the REALPHY version 1.12 method was used [17]. The whole genome of all *Actionopolyspora* sequences was submitted to the program in the Genbank format. *Salinispora tropica* was introduced as an outgroup species. The provided sequences were mapped to each other via bowtie2 aligning tool [18]. The sequence alignments of phylogeny were performed using PhyML. The phylogenetic tree was edited using MEGA-6 program [19].

3. Results and Discussion

A detailed summary of the genome features of *Act. iraqiensis* IQ-H1 is shown in **Table 2**. From the results, it can be seen that the genome size of *Act. iraqiensis* IQ-H1 was around 4.0 Mpb and the number of contigs was 110 with the largest one 217989 pb. The data obtained from the whole genome sequences have shown that the genome sizes of *Actinopolyspora* genus ranges from 4.23 Mpb as in the case of *Act. mortivallis* to 5.25 Mpb as in case of *Act. halophila* [20, 21]. The results have also shown that the GC content of *Act. iraqiensis* IQ-H1 was 70.46%. Previous studies found that Gram-positive bacteria have high GC content than Gram-negative bacteria [22]; also found that GC content is positively correlated with the genome size in bacteria [23]. Although *Act. iraqiensis* IQ-H1 has the smallest genome size compared to the other species of *Actinopolyspora* that are included in this study (**Table 1**), it can be seen that *Act. iraqiensis* IQ-H1 has the highest GC content. However, this is clearly because that the whole genome of *Act. iraqiensis* IQ-H1 was not sequenced completely as the genome sequence of this strain was deposited as a draft genome with 110 contigs and many regions from the genome might be missing. However, only the *Act. erythraea* YIM 90600 genome of the genus *Actinopolyspora* was sequenced completely [21]. Other genome features, including protein coding genes (CDS), tRNA genes, rRNA genes, open reading frames (ORF) and GC skew as well as GC content, are shown in **Figure 1**.

For more information about the Conference please visit the websites:

<http://ihicps.com/>

Table 2. Genome features of *Act. iraqiensis* IQ-H1 draft genome.

Attribute	Value
Genome total length	3827684
Contigs	110
G+C content (%)	70.46
Largest contig	217989
Protein coding genes (CDS)	4039
tRNA genes	53
rRNA genes (23S, 16S, 5S)	5
N50	87719
L50	146

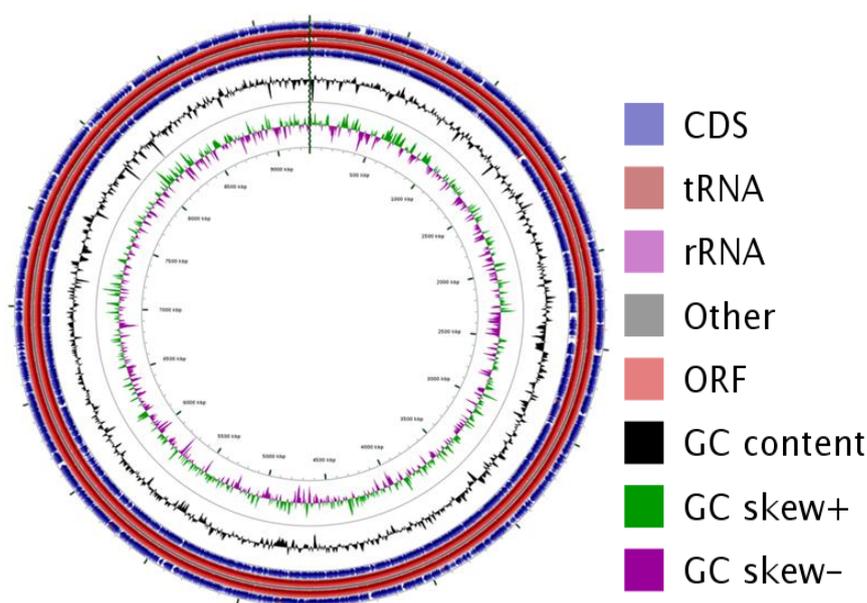


Figure 1. Circular genome representation of *Act. iraqiensis* IQ-H1 draft genome. The inner most ring represents chromosome position megabase pairs. The next rings represent the feature regions which are indicated in different colors..

It was necessary for determining the core and the pan-genome of the candidate *Actinopolyspora* species to identify genomic features common to all and to distinguish those that are unique to *Act. iraqiensis* IQ-H1. The results from the core and pan-genome analysis reveal that the pan-genome size of the nine strains of *Actinopolyspora* comprises of 15 Mpb (**Figure 2**). However, studies have shown that adding more bacterial genome sequences result in an expansion in the pan-genome size of a bacterial species which is known as open pan-genome [24, 25]. The outer-most slot (red) represents the core genome of *Act. iraqiensis* IQ-H. The core genome of *Act. iraqiensis* IQ-H1 is estimated to be 2.2 Mpb (from 12.8-15 Mpb, **Figure 2**). The core genome slot shows regions, where a BLAST hit, was present between the reference sequence, *Act. erythraea* YIM 90600 in this case, and all of the other genome sequences. Moreover, rapid advances in whole-genome sequencing methodologies have led to an enormous increase in the number of bacterial genomes deposited Genbank databases [26].

For more information about the Conference please visit the websites:

<http://ihicps.com/>

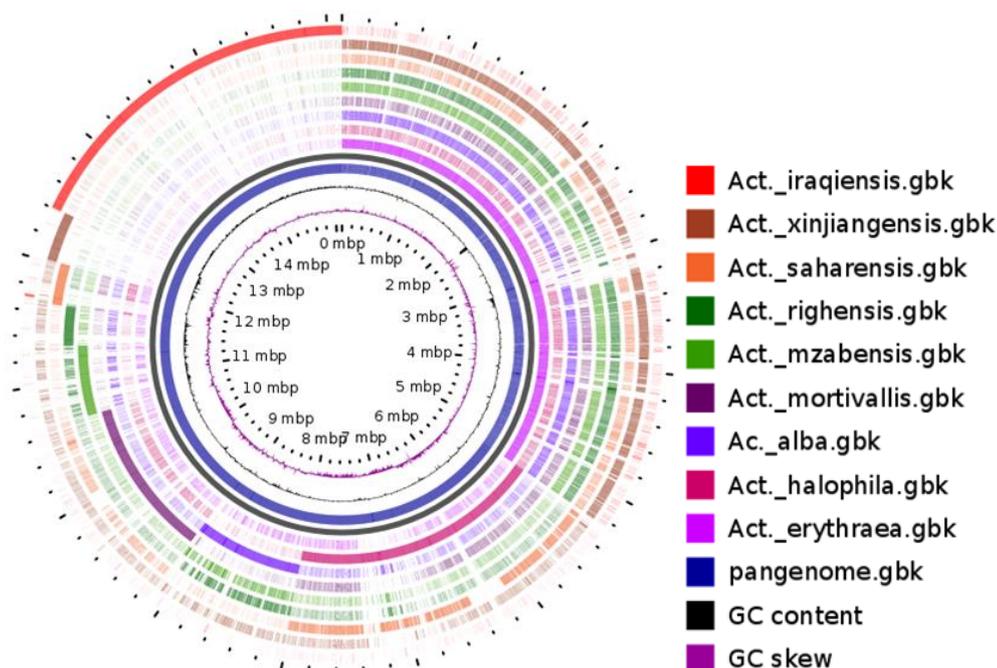


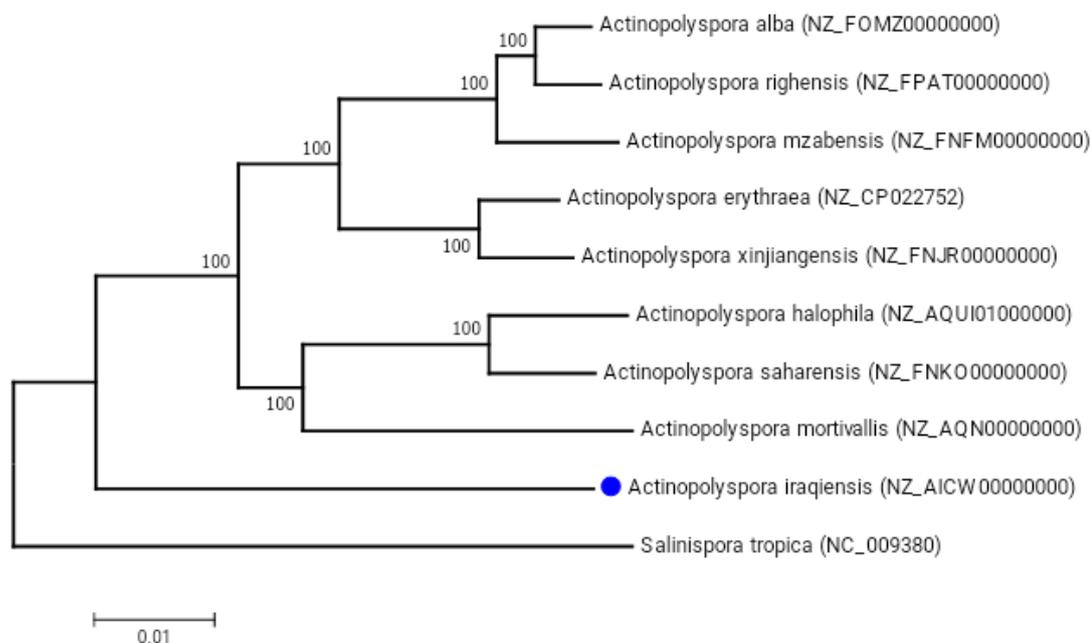
Figure 2. Core and pan-genome analysis map using the nine *Actinopolyspora* genomes. The inner-most slot (blue) shows the constructed pan-genome using all uploaded genome sequences. The next slots show regions where there were BLAST hits between the constructed pan-genome and the other uploaded genomes. The large gaps show regions missing from the seed genome (*Act. erythraea* YIM 90600) but found in one of the other genomes. These regions were appended onto the reference pangenome and are thus visible as gaps in the BLAST results for the seed genome slot (*Act. erythraea* YIM 90600).

In the era of genome sequencing and bioinformatics, it is now generally accepted that genome sequencing has the potential to be a routine approach of measuring genetic relatedness between closely related species. It has been demonstrated in many studies, the analysis of tetranucleotide usage patterns is often as a much more reliable measure of sequence relatedness than the GC content the traditional method of DNA-DNA hybridisation [15, 27, 28]. The threshold value > 0.999 (Above cut-off) indicates that the two genomes are the same species while the threshold value > 0.989 (Below cut-off) indicates that the two genomes are distinctly different [15]. Based on these values, the results have shown that *Act. iraqiensis* IQ-H1 is a new distinct species in the genus of *Actinopolyspora* (**Table 3**). However, *Act. halophila* DSM 43834, *Act. mortivallis* DSM 44261 and *Act. saharensis* DSM 45459 seem to be the most relative strains to *Act. iraqiensis* IQ-H1 with z-scores 0.99784, 0.98943 and 0.99789 respectively. Remarkably, it was noticed that when two genomes are closely related the distinction between the z-scores values will decrease while when the relatedness between two genomes decreased, the disparity between the z-scores values will increased [29]. According to this, *Act. iraqiensis* is the most related species to *Act. saharensis* DSM 45459 and *Act. iraqiensis* is the most distinct species to *Act. righensis* DSM 45501 with z-scores 0.99789 and 0.89345 respectively (**Table 3**).

Table 3. Tetranucleotide signature results for the relatedness between *Act. iraqiensis* IQ-H1 genome and the related species in the genus *Actinopolyspora* based on whole/draft genome sequences.

Genome sequences		Z-Score		
Query Genome	Reference Genomes	Above cut-off (> 0.999)	In range (> 0.989)	Below cut-off (< 0.989)
<i>Act. iraqiensis</i> IQ-H1	<i>Act. halophila</i> DSM 43834		0.99784	
	<i>Act. mortivallis</i> DSM 44261		0.98943	
	<i>Act. erythraea</i> YIM 90600			0.92642
	<i>Act. righensis</i> DSM 45501			0.89345
	<i>Act. alba</i> DSM 45004			0.89534
	<i>Act. saharensis</i> DSM 45459		0.99789	
	<i>Act. mzabensis</i> DSM 45460			0.89651
	<i>Act. xinjiangensis</i> DSM 46732			0.9182

It was observed that using the whole genome sequence for phylogenetic analysis is quite complicated and that the phylogenetic trees based on whole-genome analysis are not similar [30]. In our study, we utilized a REALPHY bioinformatics program [17] to infer a phylogenetic tree of *Act. iraqiensis* IQ-H1 with those of closely related *Actinopolyspora* strains based on whole genome sequences. From the results, it is very obvious that *Act. iraqiensis* IQ-H1 is the most closely related to *Act. saharensis* DSM 45459, *Act. halophila* DSM 43834 and *Act. mortivallis* DSM 44261 (**Figure 3**).

**Figure 3.** A maximum likelihood phylogenetic tree constructed from the nine whole genome sequences in the GenBank format using the REALPHY method [17] via bowtie2 aligning tool [18]. *Salinispora tropica* was introduced as an outgroup species. The phylogenetic tree was edited using MEGA-6 program [19]. The

For more information about the Conference please visit the websites:

<http://ihicps.com/>

phylogeny was tested by 1000 of bootstrap replications.

Figure 3. A maximum-likelihood phylogenetic tree constructed from the nine whole genome sequences in the GenBank format using the REALPHY method [17] via bowtie2 aligning tool [18]. *Salinispora tropica* was introduced as an outgroup species. The phylogenetic tree was edited using MEGA-6 program [19]. The phylogeny was tested by 1000 of bootstrap replications.

4. Conclusion

This study has shown that *Actinopolyspora iraqiensis* IQ-H1 is an Iraqi novel strain of actinobacteria. *Act. iraqiensis* IQ-H1 was closely related strain to *Act. halophila* DSM 43834, *Act. mortivallis* DSM 44261 and *Act. saharensis* DSM 45459 based on z-scores; and most related strain to *Act. saharensis* DSM 45459, *Act. halophila* DSM 43834 and *Act. mortivallis* DSM 44261 based on whole genome sequences phylogenetic tree. The findings indicate that the biological information in the form of whole genome sequences stored at the National Center for Biotechnology Information (NCBI) database along with the bioinformatics tools used in the study can be utilized for the molecular phylogenetics and genomic features analyses of *Act. iraqiensis* IQ-H1 and related species of the genus *Actinopolyspora*.

References

1. Gochnauer, M.; Leppard, G.; Komaratat, P.; Kates, M.; Novitsky, T.; Kushner, D. Isolation and characterization of *Actinopolyspora halophila*, gen. et sp. nov., an extremely halophilic actinomycete. *Can. J. Microbiol.* **1975**, *21*, 1500-1511, doi: 10.1139/m75-222.
2. Zhi, X.Y.; Li, W.J.; Stackebrandt, E. An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class Actinobacteria, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. *Int. J. Syst. Evolut. Microbiol.* **2009**, *59*, 589-608, doi: 10.1099/ijs.0.65780-0.
3. Parte, A.C. LPSN-List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. *Int. J. Syst. Evol. Microbiol.* **2018**, *68*, 1825-1829, doi: 10.1093/nar/gkt1111.
4. Zhi, X.; Yang, L.; Wu, J.; Tang, S.; Li, W. Multiplex specific PCR for identification of the genera *Actinopolyspora* and *Streptomonospora*, two groups of strictly halophilic filamentous actinomycetes. *Extremophiles.* **2007**, *11*, 543-548, doi.org/10.1007/s00792-007-0066-1
5. Newman, D.J.; Cragg, G.M.; Snader, K.M. Natural products as sources of new drugs over the period 1981– 2002. *J. Nat. Prod.* **2003**, *66*, 1022-1037, doi.org/10.1021/np030096l.
6. Benhadj, M.; Gacemi-Kirane, D.; Menasria, T.; Guebla, K.; Ahmane, Z. Screening of rare actinomycetes isolated from natural wetland ecosystem (Fetzara Lake, northeastern Algeria) for hydrolytic enzymes and antimicrobial activities. *J. King Saud University-Sci.* **2019**, *31*, 706-712, doi.org/10.1016/j.jksus.2018.03.008.
7. Johnson, K. G.; Lanthier, P.H. β -Lactamases from *Actinopolyspora halophila*, an extremely halophilic actinomycete. *Arch. Microbiol.* **1986**, *143*, 379-386, doi.org/10.1007/BF00412806.
8. Yoshida, M.; Matsubara, K.; Kudo, T.; Horikoshi, K. *Actinopolyspora mortivallis* sp. nov., a moderately halophilic actinomycete. *Int. J. Syst. Evol. Microbiol.* **1991**, *41*, 15-20, doi.org/10.1099/00207713-41-1-15.
9. Ruan, J.S.; Al-Tai, A.M.; Zhou, Z.H.; Qu, L.H. *Actinopolyspora iraqiensis* sp. nov., a new halophilic actinomycete isolated from soil. *Int. J. Syst. Evol. Microbiol.* **1994**, *44*, 759-763, doi.org/10.1099/00207713-44-4-759.

For more information about the Conference please visit the websites:

<http://ihicps.com/>

10. Aziz, R.K.; Bartels, D.; Best, A.A.; DeJongh, M.; Disz, T.; Edwards, R. A.; Formsma, K.; Gerdes, S.; Glass, E. M.; Kubal, M.; Meyer, F. The RAST Server: rapid annotations using subsystems technology. *BMC genomics*. **2008**, *9*, 75, doi.org/10.1186/1471-2164-9-75.
11. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. **2013**, *29*, 1072-1075, doi: 10.1093/bioinformatics/btt086.
12. Lowe, T. M.; Chan, P. P. tRNAscan-SE On-line: Search and Contextual Analysis of Transfer RNA Genes. *Nucl. Acids Res*. **2016**, *44*, W54-57, doi: 10.1093/nar/gkw413.
13. Stothard, P.; Grant, J. R.; Van Domselaar, G. Visualizing and comparing circular genomes using the CGView family of tools. *Briefings Bioinform*. **2019**, *20*, 1576-1582, doi: 10.1093/bib/bbx081.
14. Petkau, A.; Stuart-Edwards, M.; Stothard, P.; Van Domselaar, G. Interactive microbial genome visualization with GView. *Bioinformatics*. **2010**, *26*, 3125-3126, doi: 10.1093/bioinformatics/btq588.
15. Teeling, H.; Meyerdierks, A.; Bauer, M.; Amann, R.; Glöckner, F.O. Application of tetranucleotide frequencies for the assignment of genomic fragments. *Environ. Microbiol*. **2004**, *6*, 938-947, doi: 10.1111/j.1462-2920.2004.00624.x
16. Richter, M.; Rosselló-Móra, R.; Oliver, F.; Peplies, J.J. SpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics*. **2016**, *32*, 929-931, doi: 10.1093/bioinformatics/btv681.
17. Bertels, F.; Silander, O.K.; Pachkov, M.; Rainey, P.B.; van Nimwegen, E. Automated reconstruction of whole-genome phylogenies from short-sequence reads. *Mol. Biol. Evol*. **2014**, *31*, 1077-1088, doi: 10.1093/molbev/msu088.
18. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature methods*. **2012**, *9*, 357, doi.org/10.1038/nmeth.1923.
19. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A. Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol*. **2013**, *30*, 2725-2729, doi: 10.1093/molbev/mst197.
20. Tang, S.K.; Wang, Y.; Klenk, H.P.; Shi, R.; Lou, K.; Zhang, Y. J.; Chen, C.; Ruan, J. S.; Li, W.J. *Actinopolyspora alba* sp. nov. and *Actinopolyspora erythraea* sp. nov., isolated from a salt field, and reclassification of *Actinopolyspora iraqiensis* Ruan et al. 1994 as a heterotypic synonym of *Saccharomonospora halophila*. *Int. J. Syst. Evol. Microbiol*. **2011**, *61*, 1693-1698, doi.org/10.1099/ijs.0.022319-0.
21. Chen, D.; Feng, J.; Huang, L.; Zhang, Q.; Wu, J.; Zhu, X.; Duan, Y.; Xu, Z. Identification and characterization of a new erythromycin biosynthetic gene cluster in *Actinopolyspora erythraea* YIM90600, a novel erythronolide-producing halophilic actinomycete isolated from salt field. *PLoS one*. **2014**, *9*, e108129, doi: 10.1371/journal.pone.0108129.
22. Li, X. Q.; Du, D. Variation, evolution, and correlation analysis of C+ G content and genome or chromosome size in different kingdoms and phyla. *PLoS One*. **2014**, *9*, e88339, doi.org/10.1371/journal.pone.0088339.
23. Nishida, H. Evolution of genome base composition and genome size in bacteria. *Front. Microbiol*. **2012**, *3*, 420, doi: 10.3389/fmicb.2012.00420.
24. Livingstone, P.G.; Mophew, R.M.; Whitworth, D.E. Genome sequencing and pan-Genome analysis of 23 *Coralloccoccus* spp. strains reveal unexpected diversity, with particular plasticity of predatory gene sets. *Front. Microbiol*. **2018**, *9*, 3187, doi.org/10.3389/fmicb.2018.03187.
25. Blaustein, R.A.; McFarland, A.G.; Maamar, S.B.; Lopez, A.; Castro-Wallace, S.; Hartmann, E.M. Pangenomic approach to understanding microbial adaptations within a model built environment, the International Space Station, relative to human hosts and soil. *Systems*. **2019**, *4*, e00281-18, doi: 10.1128/mSystems.00281-18.

For more information about the Conference please visit the websites:

<http://ihicps.com/>

26. Brockhurst, M.A.; Harrison, E.; Hall, J.P.; Richards, T.; McNally, A.; MacLean, C. The ecology and evolution of pangenomes. *Curr. Biol.* **2019**, *29*, R1094-R1103, doi.org/10.1016/j.cub.2019.08.012.
27. Liu, Y.; Yang, D.; Zhang, N.; Chen, L.; Cui, Z.; Shen, Q.; Zhang, R. Characterization of uncultured genome fragment from soil metagenomic library exposed rare mismatch of internal tetranucleotide frequency. *Front. Microbiol.* **2016**, *7*, 2081, doi.org/10.3389/fmicb.2016.02081.
28. Siranosian, B.; Perera, S.; Williams, E.; Ye, C.; de Graffenried, C.; Shank, P. Tetranucleotide usage highlights genomic heterogeneity among mycobacteriophages. *F1000Research.* **2015**, *4*, doi: 10.12688/f1000research.6077.2.
29. Richter, M.; Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Nat. Acad. Sci.* **2009**, *106*, 19126-19131, doi.org/10.1073/pnas.0906412106.
30. Dutilh, B.E.; Huynen, M.A.; Bruno, W.J.; Snel, B. The consistent phylogenetic signal in genome trees revealed by reducing the impact of noise. *J. Mol. Evol.* **2004**, *58*, 527-539, doi.org/10.1007/s00239-003-2575-6.
31. Auch, A.; von Jan, M.; Klenk, H.P.; Göker, M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Standards in Genomic Sciences.* **2010**, *2*, 117-134, doi: 10.4056/sigs.531120
32. Konstantinidis, K.; Tiedje, M. Genomic insights that advance the species definition for prokaryotes. *Proc. Nat. Acad. Sci. United States Am.* **2005**, *102*, 2567–2572, doi: 10.1073/pnas.0409727102.
33. Varghese, N.; Mukherjee, S.; Ivanova, N.; Konstantinidis, K.T.; Mavrommatis, K.; Kyrpides, N.C.; Pati, A. Microbial species delineation using whole genome sequences. *Nucl. Acids Res.* **2015**, *43*, 6761–6771, doi: 10.1093/nar/gkv657.
34. Jain, C.; Rodriguez-R, L.M.; Phillippy, A.M.; Konstantinidis, K.T.; Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **2018**, *9*, 5114, doi.org/10.1038/s41467-018-07641-9.