

COMMENTARY

Open Access



MixS-HCR: a MixS extension defining a minimal information standard for sequence data from environments pertaining to hydrocarbon resources

Nicolas Tsesmetzis^{1*} , Pelin Yilmaz², Peter C. Marks¹, Nikos C. Kyrpides³, Ian M. Head⁴ and Bart P. Lomans⁵

Abstract

Here we introduce a MixS extension to facilitate the recording and cataloguing of metadata from samples related to hydrocarbon resources. The proposed MixS-HCR package incorporates the core features of the MixS standard for marker gene (MIMARKS) and metagenomic (MIMS) sequences along with a hydrocarbon resources customized environmental package. Adoption of the MixS-HCR standard will enable the comparison and better contextualization of investigations related to hydrocarbon rich environments. The insights from such standardized way of reporting could be highly beneficial for the successful development and optimization of hydrocarbon recovery processes and management of microbiological issues in petroleum production systems.

Keywords: Hydrocarbon resources, Sequence data, MixS standards, Environmental package

Background

Hydrocarbon Occurrences are defined as the natural and artificial environmental features that are rich in hydrocarbons (Fig. 1). Hydrocarbon Occurrences that can be exploited in a commercially viable manner are designated as Hydrocarbon Resources (dotted frame in Fig. 1). HCRs currently cover over 80 % of our global energy needs and they will continue to do so through 2040 [1]. Contrary to the public perception, these hydrocarbon-rich environments are often inhabited by microorganisms. In situ processes including oil degradation, methane generation, and hydrogen sulfide production are to a great extent driven or accelerated by the activity of microorganisms in these systems [2]. Moreover, microorganisms are also implicated in metal corrosion and fouling of the hydrocarbon production and transport infrastructure [2]. If left uncontrolled, such microbial (or microbially influenced) processes can lead to adverse environmental and operational consequences. On the other hand, novel

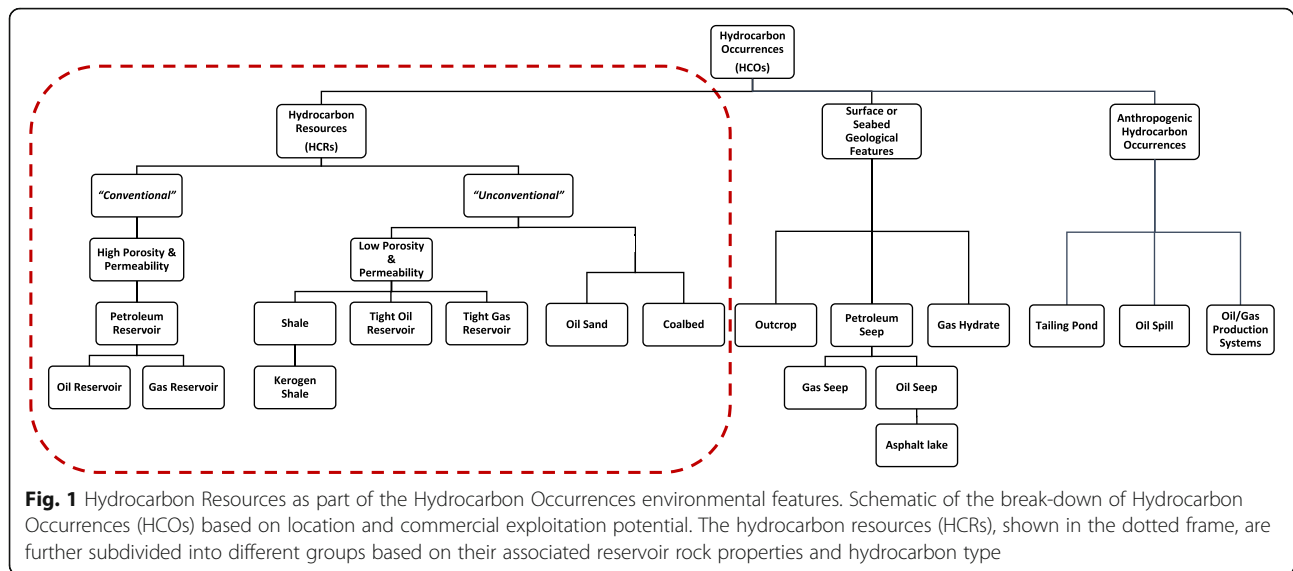
applications where microbes can play a more positive role in petroleum systems are becoming increasingly evident. These include hydrocarbon exploration, reservoir souring prevention, hydrocarbon upgrading and enhanced hydrocarbon and energy recovery.

Restraining or harnessing the potential of these microbes requires a considerable understanding of their metabolisms and role in these environments. Tools and methodologies that allow the enumeration, taxonomic identification, and metabolic prediction of these microbes are therefore essential towards this goal. Such methodologies will enable us to address questions relating to microbial spatiotemporal variability and what drives them within an HCR. This will also help us determine whether particular microorganisms were present in the HCR at the time of deposition or whether they were introduced at a later stage through fluid migration (e.g. during the formation or recovery of the hydrocarbon resource). Other interesting questions pertain to hydrocarbon recovery processes, and especially the effect of water injection (aka “water flooding”) on the indigenous and introduced microbial communities associated with HCRs. One particular operational issue related to this is stimulation of biogenic H₂S production, an undesirable

* Correspondence: nicolas.tsesmetzis@shell.com

¹Shell International Exploration and Production Inc., 3333 HW6S, Houston 77082, TX, USA

Full list of author information is available at the end of the article



process better known as reservoir “souring”. Moreover, in terms of microbial monitoring, such methodologies will help us determine to what extent the indigenous microbial community of an HCR is represented by the analysis of readily recoverable produced fluids, and if analysis of produced fluids can reduce the need to analyze material directly recovered from the hydrocarbon rich formation, such as reservoir cores which are much more difficult to obtain. It will also be practically important if the analysis of planktonic populations from produced fluids can provide an alert to potential MIC phenomena caused by sessile populations. A growing inventory of HCR systems will also allow us to determine if specific beneficial or detrimental microbially catalyzed processes in HCRs are the result of the same microbial populations in most instances, or whether a particular process is driven by different organisms in different HCR environments.

In addition to environmental parameters, the outcome from microbiological surveys of HCRs can be strongly impacted by logistical and technical constraints. These include sample acquisition, transport and preservation; the choice of DNA amplification primers (i.e. coverage, specificity, target gene); enumeration methods (i.e. MPN vs. ATP vs. qPCR vs. Phylochip vs. Next-Gen sequencing, etc.); sequencing platform (Sanger, 454, IonTorrent, Illumina, etc.) and related downstream bioinformatics pipelines.

Need for standards

There is a growing list of studies on HCRs, but their collective retrieval (or their corresponding sequences from public repositories) is currently impossible even with complex keyword searches. Moreover, the majority of available datasets lack sufficient contextual data, which would facilitate more comprehensive comparative analysis [3–6].

In order to maximize the knowledge gained from these largely unexplored microbial ecosystems it is important to formalize and standardize environment descriptors for studies of these habitats. It is equally important to define a minimum set of contextual parameters, which should accompany the submission of sequence information from HCR studies to the International Nucleotide Sequence Database Collaboration [7]. The adoption of such standardization would drastically improve the quality, accessibility and value of the HCR-related information residing in INSDC.

The need for standardization is not novel or unique to HCR studies. Since 2005, the Genomic Standards Consortium [3, 4] has made remarkable efforts in proposing standards for genomic (MIGS) and metagenomic (MIMS) sequences [8, 9] as well as for marker genes (MIMARKS) [10] and biosynthetic gene clusters (MIBiG) [11]. Moreover, a single entry point to all the minimum information specifications was also proposed (Minimum Information about any Sequence; MIxS) [10]. Equally important, the GSC proposed a wide range of environmental packages, which cover a broad range of the commonly encountered environments in research studies (i.e. human-associated, soil, water, sediments, built environments, etc.) [10, 12].

Around the same time a separate effort aimed at the development of ENVO, a standardized and semantically controlled representation of environment descriptors was undertaken [13]. ENVO quickly became a core component of the MIxS specification. Despite being extensively developed for other environmental features and habitats, ENVO currently has very limited content related to HCR. For example, in the MIxS specification under the ENVO term *biome* only the subclasses *aquatic biome*, *polar biome*, and *terrestrial biome* are currently present. The term *subterranean biome* (or *subterrestrial biome*), which would include biomes related to certain HCR (e.g.

hydrocarbon reservoir) as well as other subterranean biomes [14–16] is currently missing. Similarly, in the case of ENVO's *environmental feature* branch, additional HCR terms such as *gas reservoir*, *oil sand*, and *coalbed* will need to be included in addition to the existing *oil reservoir* term. Finally, in the *environmental material* section of ENVO, HCR terms like *formation water*, *injection water*, *drilling fluid*, *tailing pond* and many more would supplement the existing *oil field production water* term. It is therefore apparent that expansion of ENVO to include HCR-related terms will greatly benefit the standardization of a growing number of studies on these environments. An initiative to introduce such HCR-related terms in ENVO is currently underway.

Implementation of a Hydrocarbon Resource Environmental package

In an effort to assist with the standardization of data acquisition and observations derived from HCR-related environments we introduce the MlXS-HCR minimum information standard. This standard is tailored for HCR-related studies and aims at capturing key environmental parameters influencing microbial activity in these environments and standardizing their method of reporting. This is accomplished by the adoption of terms (such as temperature, pressure, porosity, etc.) from previously reported environmental packages (i.e. Water, Sediment, Wastewater/Sludge, etc.) as well as the introduction of new checklist items specific to these environments. A checklist consisting of 93 fields from several disciplines including geology, geochemistry, petrophysics, reservoir engineering, and production chemistry has been compiled (Table 1 and [MlXS HCR detailed table] in Additional file 1). Some of the included terms pertain to the HCR entity as a whole whereas others concern the sample(s) acquired from that entity. Moreover, the checklist is divided into 5 sections to facilitate the grouping of items derived from the same type of analysis or the same topic (Table 1). These sections include general information about the HCR, descriptors related to the HCR's production history, the sample's hydrocarbon and water chemistry, sampling procedures, sample transport and storage conditions.

Amongst the different minimal information standards mentioned above, the MIMS and MIMARKS survey sequence specifications are probably the most relevant ones for HCR-related studies as the majority of these studies involve single gene (i.e. 16S rRNA, *dsr*, *nar*, etc.) or whole metagenome surveys. As such, in addition to the HCR environmental package, the MlXS-HCR extension also includes a subset of the MlXS checklist containing MIMS and/or MIMARKS survey fields (depending on the study) (Additional file 1). This newly proposed MlXS-HCR minimum information standard provides the foundation for

consistent capture and reporting of valuable contextual (i.e. environmental, biological and technical) information derived from HCR-related studies. An example of a MlXS-HCR-compliant report from a Brunei oil field is included in Additional file 1 [see MlXS HCR detailed table].

Development process & research community

The need for standardization of HCR-related biological information has been the topic of several conference and workshops discussions where both academia and industry acknowledged the importance of adhering to standardized ways of sharing and reporting information. MlXS-HCR minimum information standard is the joint effort of a multidisciplinary community from academia and industry including the GSC MlXS developers, environmental microbiologists, bioinformaticians, geochemists, reservoir engineers, production chemists and computer scientists. During its development, the proposed HCR environmental package sought feedback and endorsement from researchers in academia and industry working in this scientific field. A web forum was set up to promote the development and refinement of the package as well as stimulate discussion around the topic and its content [17]. Changes to the package were subject to the consensus-based agreement amongst the researchers involved in this effort. The continuous contribution to the web forum and the adoption of this standard by the research community are key elements for the success of this initiative. Like its first release, which has already gained approval by the GSC board, yearly reviews of the MlXS-HCR standard performed by the MlXS-HCR web forum coordinators [18] will be incorporated in the next available MlXS public release following review and approval by the GSC board. As with all GSC projects, news and updates pertaining to this standard are managed via the corresponding project page on the GSC website [19] but also through the MlXS-HCR web forum [17]. The latest GSC-approved downloadable version of the MlXS-HCR extension is available under the GSC MlXS extensions webpage where additional information such as a list of terms, contact details and project information are also provided [20]. Each field in the supplied spreadsheet is accompanied by a definition, an expected value (including controlled vocabulary terms where applicable), the number of occurrences each field may be used, a value syntax, a preferred unit (if applicable) and other relevant recommendations (see also [MlXS HCR detailed table] in Additional file 1).

Endorsement of this MlXS-HCR minimum information standard by the GSC, strengthens the case for its incorporation by the INSDC in the list of prerequisites at the time of sequence submission. The MlXS-HCR minimum set will be complemented with other minimum sets (currently under development) describing hydrocarbon-rich

Table 1 MlXS-HCR package items. More detailed description of the field names is provided in the Additional file 1 [see MlXS HCR detailed table]

MlXS-HCR package
HCR general properties
Hydrocarbon resource type
Hydrocarbon type produced
Basin name
Field name
Reservoir name
Hydrocarbon resource original temperature
Depth (TVDSS) of hydrocarbon resource temperature
Hydrocarbon resource original pressure
Depth (TVDSS) of hydrocarbon resource pressure
Permeability
Porosity
Lithology
Depositional environment
Hydrocarbon resource geologic age
Oil water contact depth
Formation water salinity
Sulfate in formation water
Vfa in formation water
Source rock kerogen type
Source rock lithology
Source rock depositional environment
Source rock geologic age
HCR production properties
Production start date
Production rate
Water production rate
Water cut
Injection water fraction
Secondary and tertiary recovery methods and start date
Injection water breakthrough date of specific well
Biocide administration
Biocide administration method
Chemical treatment
Chemical treatment method
Corrosion rate at sample location
HCR sampling properties
Sample well name
Well identification number
Sample type
Sample subtype
Sample collection point

Table 1 MlXS-HCR package items. More detailed description of the field names is provided in the Additional file 1 [see MlXS HCR detailed table] (*Continued*)

Temperature
Pressure
Sample true vertical depth subsea
Sample measured depth
Elevation
Oxygenation status of sample
Preservative added to sample
Sample transport conditions
Sample storage temperature
Sample storage duration
Sample storage location
Sample vol., weight, or area swabbed for DNA extraction
Organism count
Organism count qPCR information
HCR water chemistry
pH
Sample salinity
Alkalinity
Sulfate
Sulfide
Total sulfur
Nitrate
Nitrite
Ammonium
Total nitrogen
Dissolved iron
Sodium
Chloride
Potassium
Magnesium
Calcium
Total iron
Dissolved organic carbon
Dissolved inorganic carbon
Dissolved inorganic phosphorus
Total phosphorus
Suspended solids
Density
Dissolved carbon dioxide
Dissolved oxygen
Volatile fatty acids
Benzene
Toluene

Table 1 MlxS-HCR package items. More detailed description of the field names is provided in the Additional file 1 [see MlxS HCR detailed table] (*Continued*)

Ethylbenzene
Xylene
HCR hydrocarbon chemistry
API gravity
Total Acid Number
Viscosity
Pour point
Saturates wt%
Aromatics wt%
Resins wt%
Asphaltenes wt%
Other properties
Miscellaneous parameter
Additional info

environments which are not covered under the HCR term. These include anthropogenic hydrocarbon occurrences (e.g. oil and gas production systems) as well as surface and seabed hydrocarbon occurrences (e.g. cold seeps, outcrops, gas hydrates, etc.) (Fig. 1). Many of the MlxS-HCR fields are going to be shared across the different hydrocarbon occurrence types whereas new ones, specific for each of the other types, will also be proposed. Of particular importance will be the development of minimal information standards for sequence data from oil and gas production systems as these systems are allegedly subject to failures frequently attributed to MIC [21] raising environmental, safety and operational concerns.

Conclusions

The newly proposed MlxS-HCR minimum information standard provides the foundation for consistent capture and reporting of valuable contextual information derived from studies pertained to hydrocarbon resources. Its first release has already gained approval by the GSC board and will be incorporated in INSDC's sequence submission process. A web forum has also been set up to promote MlxS-HCR future improvements and extension to cover a wider range of hydrocarbon occurrences. Active involvement of the research community and adoption of the MlxS-HCR standard are key elements to the success of this initiative.

Additional file

Additional file 1: MlxS HCR detailed table. Detailed description of the MlxS-HCR fields including terms, definitions, field requirements, syntax and examples. (XLSX 43 kb)

Abbreviations

AHO: Anthropogenic hydrocarbon occurrences; ATP: Adenosine triphosphate; ENVO: Environmental ontology; GSC: Genomics standards consortium; HCOs: Hydrocarbon occurrences; HCRs: Hydrocarbon resources; INSDC: International nucleotide sequence database collaboration; MIBiG: Minimal information about a biosynthetic gene cluster; MIC: Microbially influenced corrosion; MIMARKS: Minimal information about a marker gene sequence; MIMS: Minimal information about a metagenomic sequence; MlxS: Minimal information about any sequence; MPN: Most probable number; qPCR: Quantitative polymerase chain reaction; SHO: Surface and seabed hydrocarbon occurrences

Acknowledgments

The authors are thankful to the HCR community and the Genomics Standards Consortium for their valuable comments that led to the development of this minimal standard especially Jian Lu, Cor Kuijvenhoven, Andrew Bishop, Lisa Gieg, Ken Wunch, Geert van der Kraan, Brandon Morris, Heike Hoffmann, Renato De Paula and Brett Geissler for their input.

Funding

This work was partly funded by Royal Dutch Shell.

Availability of data and material

The latest MlxS-HCR specification document is available through the Genomic Standards Consortium website [22].

Authors' contributions

NT and BPL conceived the study, participated in its design and coordination and drafted the manuscript. PY, PCM, NCK and IMH participated in the design of the study and helped to draft the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Author details

¹Shell International Exploration and Production Inc., 3333 HW6S, Houston 77082, TX, USA. ²Microbial Genomics and Bioinformatics Research Group, Max Planck Institute for Marine Microbiology, Bremen, Germany. ³DOE Joint Genome Institute, Walnut Creek, CA 94598, USA. ⁴School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne NE1 7RU, UK. ⁵Shell Global Solutions International B.V., Rijswijk, Netherlands.

Received: 28 June 2016 Accepted: 5 October 2016

Published online: 12 October 2016

References

- International Energy Outlook 2013. [http://www.eia.gov/forecasts/ieo/pdf/0484\(2013\).pdf](http://www.eia.gov/forecasts/ieo/pdf/0484(2013).pdf). Accessed 27 June 2016.
- Whitby C, Skovhus TL. *Applied Microbiology and Molecular Biology in Oilfield Systems*. Netherlands: Springer Netherlands; 2011.
- Li H, Yang SZ, Mu BZ, Rong ZF, Zhang J. Molecular phylogenetic diversity of the microbial community associated with a high-temperature petroleum reservoir at an offshore oilfield. *FEMS Microbiol Ecol*. 2007;60:74–84.
- Mbadinga SM, Li KP, Zhou L, Wang LY, Yang SZ, Liu JF, et al. Analysis of alkane-dependent methanogenic community derived from production water of a high-temperature petroleum reservoir. *Appl Microbiol Biotechnol*. 2012;96:531–42.
- Wang LY, Ke WJ, Sun XB, Liu JF, Gu JD, Mu BZ. Comparison of bacterial community in aqueous and oil phases of water-flooded petroleum reservoirs using pyrosequencing and clone library approaches. *Appl Microbiol Biotechnol*. 2014;98:4209–21.
- Piubeli F, Grossman MJ, Fantinatti-Garboggini F, Durrant LR. Phylogenetic analysis of the microbial community in hypersaline petroleum produced water from the Campos Basin. *Environ Sci Pollut Res Int*. 2014;21:12006–16.

7. Nakamura Y, Cochrane G, Karsch-Mizrachi I, International Nucleotide Sequence Database C. The International Nucleotide Sequence Database Collaboration. *Nucleic Acids Res.* 2013;41:D21–4.
8. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol.* 2008;26:541–7.
9. Kottmann R, Gray T, Murphy S, Kagan L, Kravitz S, Lombardot T, et al. A standard MIGS/MIMS compliant XML Schema: toward the development of the Genomic Contextual Data Markup Language (GCDML). *OMICS.* 2008;12:115–21.
10. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, et al. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MlxS) specifications. *Nat Biotechnol.* 2011;29:415–20.
11. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, et al. Minimum Information about a Biosynthetic Gene cluster. *Nat Chem Biol.* 2015;11:625–31.
12. Glass EM, Dribinsky Y, Yilmaz P, Levin H, Van Pelt R, Wendel D, et al. MlxS-BE: a MlxS extension defining a minimum information standard for sequence data from the built environment. *ISME J.* 2014;8:1–3.
13. Buttigieg PL, Morrison N, Smith B, Mungall CJ, Lewis SE, Consortium E. The environment ontology: contextualising biological and biomedical entities. *J Biomed Semantics.* 2013;4:43.
14. Chandler DP, Brockman FJ, Bailey TJ, Fredrickson JK. Phylogenetic Diversity of Archaea and Bacteria in a Deep Subsurface Paleosol. *Microb Ecol.* 1998;36:37–50.
15. Kovacic Jr WP, Takai K, Mormile MR, McKinley JP, Brockman FJ, Fredrickson JK, Holben WE. Molecular analysis of deep subsurface Cretaceous rock indicates abundant Fe(III)- and S(0)-reducing bacteria in a sulfate-rich environment. *Environ Microbiol.* 2006;8:141–55.
16. Sahl JW, Schmidt R, Swanner ED, Mandernack KW, Templeton AS, Kieft TL, et al. Subsurface microbial diversity in deep-granitic-fracture water in Colorado. *Appl Environ Microbiol.* 2008;74:143–52.
17. MlxS-HCR web forum. <https://groups.google.com/forum/#forum/mixs-hcr>. Accessed 27 June 2016
18. MlxS-HCR web forum contact details. mixs-hcr@googlegroups.com. Accessed 27 June 2016.
19. GSC project webpage for MlxS-HCR. <http://gensc.org/projects/mixs-hcr-gsc-project/>. Accessed 27 June 2016
20. GSC MlxS extensions webpage <http://gensc.org/mixs/mixs-extensions/>. Accessed 27 June 2016
21. Vigneron A, Alsop EB, Chambers B, Lomans BP, Head IM, Tsesmetzis N. Complementary Microorganisms in Highly Corrosive Biofilms from an Offshore Oil Production Facility. *Appl Environ Microbiol.* 2016;82:2545–54.
22. Genomic Standards Consortium. <http://gensc.org/>. Accessed 27 June 2016

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

