

Standards in Genomic Sciences eJournal

<http://standardsingenomics.org>

Checklists for Authors of SIGS Short Genome Reports for *Bacteria* and *Archaea*

While data sets may continue to grow, database structures and applications change, and new goals for research emerge, a new Standards in Genomic Sciences (SIGS) eJournal will provide static, archival snapshots of data and metadata for genomes and metagenomes. Authors may use the opportunity of publishing in SIGS, the first standards-based journal in the life sciences, to make claims concerning their reported data set in terms of how it may present significant biological features, or be a new finding of a novel clade or niche. These genome and metagenome reports will serve as points of record that are enriched with interpretative commentary (such as significant biological features, or a novel clade or niche), verified by a standards-focused editorial team, and authenticated by peer review. Publication may best be pursued in coordination with sequencing centers that assist authors in providing information about genome project sequencing, assembly and annotation.

Should you be an invited author, you will receive a templated manuscript containing information about your genome sequencing project. For the sections requiring your contributed content, please consult the checklist below prior to submission of your manuscript at <http://standardsingenomics.org>. In the event that your contribution is unsolicited, please contact the editorial office (editors@standardsingenomics.org) to discuss creation of a template manuscript for the genome that you will be reporting on.

- Is the sequence you are writing about publicly available and does it have a INSDC identifier (GenBank, EMBL, DDBJ) such as CP000230? Other identifiers may also be included such as a RefSeq accession number (e.g., NC_007643).
- Is a short abstract provided (<100 words preferred; 200 words is absolute limit)?
<http://standardsingenomics.org/index.php/sigen/pages/view/abstracts>
- Is the project relevance described?
<http://standardsingenomics.org/index.php/sigen/pages/view/projectrelevance>
- Is the organism and/or genomic DNA publically available from a culture collection? If so:
 - please indicate the appropriate strain identifier or identifiers of the culture collection(s) (e.g., ATCC, DSM, CIP, CCUG, NCIMB, NCTC, LMG, JCM, NRRL, KCTC or PCC) where the organism is known to be publically available, and cite the appropriate reference(s); and
 - please also indicate whether or not the organism and genomic DNA is from a type strain of a species or subspecies that has been validly published and provide the reference. If a type strain, the organism will be available in two or more culture collections.
- Are evidence codes indicated for features such as cell shape, motility, etc that are reported in Table 1 (the organism information table). Authors are encouraged to insert footnotes concerning feature categories that are reported (IDA), cited (TAS) or inferred (NAS) as follows:
 - IDA: Inferred from Direct Assay; note, if the evidence code is IDA, then the property should have been directly observed, for the purpose of this specific publication, for a

live isolate by one of the authors, or an expert or reputable institution mentioned in the acknowledgements.

- TAS: Traceable Author Statement; i.e., a direct report exists in the literature.
- NAS: Non-traceable Author Statement; i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence.

More information on evidence codes is available at:

<http://standardsingenomics.org/index.php/sigen/pages/view/evidencecodes>

- Pre-formatted tables are provided in the template manuscript and, in general, should not be submitted as a separate file nor be appended to the end of the manuscript. Large tables may however be sent as supplemental files in CSV or MS Excel format.
 - Superscript letters should be used in table captions and table bodies to indicate information in footnotes.
 - Tables should be constructed with one data value per cell.

Table formatting of SIGS manuscripts is described in further detail at

<http://standardsingenomics.org/index.php/sigen/pages/view/tableformatting>

- Bibliography – the indexing and formatting of each citation should be done with the EndNote style available from <http://standardsingenomics.org/index.php/sigen/pages/view/bibliography>.
- Please verify the tally of chromosomes, extrachromosomal elements (plasmids), their topologies and other metadata in the template document.
- Please solicit and verify the sections of content for genome sequencing, assembly and annotation that may typically be provided by the sequencing center.
- Has a phylogenetic tree been provided?
<http://standardsingenomics.org/index.php/sigen/pages/view/trees>

To assist you, the following tasks of manuscript preparation are generally undertaken by the editorial office, typically done in coordination with your sequencing center.

- Compute paralog clusters
- Present COG counts
- Compute conserved domain counts
- Compute transmembrane domains and signal peptides
- Standard genome features and statistics (size, topology, g+c, identifiers)

Checklist for Sequencing Centers for SIGS Short Genome Reports for *Bacteria* and *Archaea*

SIGS editors are working with sequencing centers on standards-based approaches to reporting genome sequencing and annotation. During the editorial process, authors associated with genome projects are encouraged to have their sequencing centers directly contact the editorial team of the SIGS journal for feedback or further information. The “Genome sequencing information” section of a SIGS short genome report manuscript typically contains the following:

- Completion date (initial release of complete sequence and its annotation to the author, or assignation of a GenBank accession number) and data release date (i.e., assignment of an INSDC identifier)
- Genome sequencing and assembly
i.e., address coverage, quality, error rate, instrumentation, 454/Sanger, etc.
- Genome annotation
This will typically mention use of software applications such as Critica or Glimmer