

Genome analysis

JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison

Michael Richter^{1,*}, Ramon Rosselló-Móra², Frank Oliver Glöckner^{3,4} and Jörg Peplies¹

¹Ribocon GmbH, Fahrenheitstraße. 1, 28359 Bremen, Germany, ²Marine Microbiology Group, Mediterranean Institute for Advanced Studies, C/ Miquel Marqués, 21, 07190, Esporles, Spain, ³Max Planck Institute for Marine Microbiology, Microbial Genomics and Bioinformatics Research Group, Celsiusstraße 1, 28359 Bremen, Germany and ⁴Jacobs University gGmbH Bremen, Department of Life Sciences and Chemistry, Campus Ring 1, 28759 Bremen, Germany

*To whom correspondence should be addressed.

Associate Editor: John Hancock

Received on September 10, 2015; revised on October 29, 2015; accepted on November 12, 2015

Abstract

Summary: JSpecies Web Server (JSpeciesWS) is a user-friendly online service for *in silico* calculating the extent of identity between two genomes, a parameter routinely used in the process of polyphasic microbial species circumscription. The service measures the average nucleotide identity (ANI) based on BLAST+ (ANiB) and MUMmer (ANIm), as well as correlation indexes of tetra-nucleotide signatures (Tetra). In addition, it provides a Tetra Correlation Search function, which allows to rapidly compare selected genomes against a continuously updated reference database with currently about 32 000 published whole and draft genome sequences. For comparison, own genomes can be uploaded and references can be selected from the JSpeciesWS reference database. The service indicates whether two genomes share genomic identities above or below the species embracing thresholds, and serves as a fast way to allocate unknown genomes in the frame of the hitherto sequenced species.

Availability and implementation: JSpeciesWS is available at <http://jspecies.ribohost.com/jspeciesws>.

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

Contact: mrichter@ribocon.com

1 Introduction

DNA-DNA-hybridization (DDH) measures the degree of genomic similarity between two pools of DNA molecules. Despite that distinct DDH determination methods frequently lead to different results, since nearly 50 years it is considered as the gold standard for species delineation (Rosselló-Móra and Amann, 2015). In times of cost-efficient sequencing with scientists having easy access to thousands of publicly available genome sequences, the species delineation based on user friendly computational pairwise genome comparisons to determine overall genome relatedness indices

[OGRI, (Chun and Rainey, 2014)] has emerged. This approach has already been widely accepted by the scientific community (Colston *et al.*, 2014; Li *et al.*, 2015; Lugli *et al.*, 2014; Oren and Garrity, 2014) and specialized tools like the stand-alone software JSpecies (Richter and Rosselló-Móra, 2009) or similar solutions (Meier-Kolthoff *et al.*, 2013) have proven to provide valuable OGRI results.

Here we present the JSpecies Web Server (JSpeciesWS) a web-based implementation of the core features of the original JSpecies software for easy user-access and extended functionality. It represents the latest development of JSpecies in terms of usability,

flexibility, and efficiency. An online reference database of all published whole and draft genomes (about 32 000) with pre-calculated OGRI results has been added and can be searched in seconds based on genome comparison using the new Tetra Correlation Search (TCS) function.

2 Methods

JSpeciesWS has been implemented with the Google Web Toolkit (<http://www.gwtproject.org/>) version 2.7.0 and uses a Twitter-Bootstrap clone (<https://github.com/gwtbootstrap3/gwtbootstrap3>) for the user interface elements. Imported sequences get validated by using the FastaValidator version 1 (Waldmann et al., 2014).

2.1 Average nucleotide identity based on BLAST+

The average nucleotide identity based on BLAST+ (ANiB) calculation is implemented as described by Goris et al. (2007). In contrast to the original JSpecies stand-alone implementation, the web server uses BLAST version 2.2.29+ (Camacho et al., 2009) instead of BLAST version 2.2.26 (Altschul et al., 1997).

2.2 Average nucleotide identity based on MUMmer

For average nucleotide identity based on MUMmer (ANiM) calculation, the dnadiff script of the MUMmer tool version 3.0 (Kurtz et al., 2004) is used instead of a self-written parser used by the original JSpecies stand-alone implementation.

2.3 Tetra-nucleotide signature correlation index

Tetra is an alignment-free parameter correlating with ANI, and implemented in the same way as the original JSpecies stand-alone (Richter and Rosselló-Móra, 2009).

2.4 Tetra Correlation Search

TCS is a new feature of JSpeciesWS that allows rapid comparisons against the JSpeciesWS reference database, providing a list of the most similar genomes based on their resulting Tetra-nucleotide signature correlation index.

2.5 Genomes reference database

The internal JSpeciesWS genome database resembles NCBI's genomic sequence data taken from <ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank> and includes all primary submissions of assembled genome sequences and their associated annotation data. Currently, more than 32 000 prokaryotic genomes at different stages of finishing are included. For each entry, taxonomic information is provided for the field's domain, phylum, class, order, family, genus and species. Further contextual data extracted from the original entry (if available) include genome size, guanine-cytosine content, Gram staining, shape, arrangement, endospores formation, motility, salinity, oxygen, habitat and temperature range.

3 Usage

JSpeciesWS does not require user registration. By visiting the website, a session is generated linked to a dedicated genome cart. The genome cart can be filled with own genome uploads and/or genomes can be added from the JSpeciesWS reference database. The cart (session) expires automatically after a period of 14 days of inactivity (no

revisit based on session code provided) and all data will be deleted. The limit of genomes per cart is currently set to 15.

Pairwise ANiB, ANiM or Tetra calculations can be performed between all genomes in the cart (all vs all) or, alternatively, only one genome of the cart can be selected for comparison against all other genomes in the cart to reduce calculation time. In addition, a TCS can be performed for each genome in the cart against the complete JSpeciesWS reference database of all published whole and draft genomes.

After starting an analysis, the status field is regularly updated and the user is finally informed about a finished calculation by a pop-up message. In addition, the user can provide an e-mail address to become informed about the completion of the calculations. A all versus all comparison of 15 genomes (210 calculations) takes on the current system about 2 h based on ANiB, 30 min based on ANiM and 10 min based on Tetra. All results can be revisited, exported for further processing and shared among users.

4 Conclusion

Recent studies have shown that electronic (*in silico*) pairwise OGRI can replace standard laboratory-based DDH in order to simplify, and at the same time substantiate, this central aspect of prokaryotic species delineation. In addition, OGRI allows taxonomists to quickly verify the species affiliation of any public genome sequence and to identify wrongly annotated submissions. JSpeciesWS will help to further propagate this approach by providing fast and easy access to the corresponding analyses and a comprehensive genomes reference database.

Funding

This research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 311975. This publication reflects the views only of the author, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

Conflict of Interest: none declared.

References

- Altschul, S.F. et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Camacho, C. et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
- Chun, J. and Rainey, F.A. (2014) Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. *Int. J. Syst. Evol. Microbiol.*, **64**, 316–324.
- Colston, S.M. et al. (2014) Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using aeromonas as a test case. *mBio*, **5**, e02136.
- Goris, J. et al. (2007) DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.*, **57**, 81–91.
- Kurtz, S. et al. (2004) Versatile and open software for comparing large genomes. *Genome Biol.*, **5**, R12.
- Li, X. et al. (2015) The relationship of the whole genome sequence identity to DNA hybridization varies between genera of prokaryotes. *Antonie Van Leeuwenhoek*, **107**, 241–249.
- Lugli, G.A. et al. (2014) Investigation of the evolutionary development of the genus bifidobacterium by comparative genomics. *Appl. Environ. Microbiol.*, **80**, 6383–6394.

- Meier-Kolthoff, J.P. *et al.* (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*, **14**, 60.
- Oren, A. and Garrity, G.M. (2014) Then and now: a systematic review of the systematics of prokaryotes in the last 80 years. *Antonie Van Leeuwenhoek*, **106**, 43–56.
- Richter, M. and Rosselló-Móra, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. USA*, **106**, 19126–19131.
- Rosselló-Móra, R. and Amann, R. (2015) Past and future species definitions for Bacteria and Archaea. *Syst. Appl. Microbiol.*, **38**, 209–216.
- Waldmann, J. *et al.* (2014) FastaValidator: an open-source Java library to parse and validate FASTA formatted sequences. *BMC Res. Notes*, **7**, 365.