Bioinformatics, YYYY, 0–0
doi: 10.1093/bioinformatics/xxxxx
Advance Access Publication Date: DD Month YYYY
Manuscript Category

Genome Analysis

PLACNETw: a web-based tool for plasmid reconstruction from bacterial genomes

Luis Vielva[1], María de Toro[2], Val F. Lanza[3] and Fernando de la Cruz[4*]

[1] Departamento de Ingeniería de Comunicaciones, Universidad de Cantabria, Santander, Spain, [2] Plataforma de Genómica y Bioinformática, Centro de Investigación Biomédica de La Rioja (CIBIR), Logroño, Spain [3] Departamento de Microbiología, Instituto de Investigación Sanitaria Ramón y Cajal (IRYCIS), Madrid, Spain, [4] Departamento de Biología Molecular, Universidad de Cantabria and Instituto de Biomedicina y Biotecnología de Cantabria (Universidad de Cantabria- CSIC), Santander, Spain.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

PLACNET is a graph-based tool for reconstruction of plasmids from next generation sequence pair-end datasets. PLACNET graphs contain two types of nodes (assembled contigs and reference genomes) and two types of edges (scaffold links and homology to references). Manual pruning of the graphs is a necessary requirement in PLACNET, but this is difficult for users without solid bioinformatic background. PLACNETw, a web tool based on PLACNET, provides an interactive graphic interface, automates BLAST searches, and extracts the relevant information for decision making. It allows a non-trained user to visualize the assembly graphs and related information of contigs as well as reference sequences, so that the pruning operations can be done interactively from a personal computer without the need for additional tools. After pruning, each plasmid becomes a separate connected component subgraph. The resulting data is automatically downloaded by the user.

Availability: PLACNETw is freely available at https://castillo.dicom.unican.es/upload/ .

Contact: LuisVielva@gmail.com

Supplementary information: A tutorial video and various solved examples are available at Bioinformatics online and at the PLACNETw main page.

Introduction

Horizontal Gene Transfer (HGT) is a fundamental driver of bacterial evolution. It is, in many cases, responsible for the acquisition of antimicrobial resistance and virulence determinants. Plasmids are main players in HGT events, and thus essential constituents of bacterial genomes (Wiedenbeck and Cohan, 2011). Characterization of plasmids in whole genome sequence (WGS) data is hindered by methodological limitations. Short-read sequencing cannot resolve repeated elements, resulting in hundreds of contigs per genome (Treangen and Salzberg, 2011). PlasmidFinder (Carattoli *et al.*, 2014) and ACLAME (Leplae *et al.*, 2010) are tools developed to assist in plasmid analysis. While both can identify and analyze individual plasmid contigs, neither was designed to assemble them. Besides, PlasmidSPAdes (Antipov *et al.*, 2016) enables the assembly of plasmid contigs, mainly by exploiting differences in coverage. We recently published PLACNET, a graph-based method to identify, visualize and analyze plasmids in WGS projects by creating a network of contig interactions (Lanza *et al.*, 2014). PLACNET allows the identification of plasmids, providing a useful tool to carry out comprehensive plasmid population genetic studies. Compared with previous methods, which are automated (Arredondo-Alonso *et al.*, 2017), PLACNET relies on manual tinkering (pruning) the graph, which results in added precision and sensitivity (de Been *et al.*, 2014). Nevertheless, using PLACNET was a difficult task for users without solid bioinformatic background. To address this issue, this work develops PLACNETw, a web-based tool that allows scientists with no specific bioinformatic background, or low computational resources, to explore plasmids in WGS data. Of course, optimal network pruning relies on a user with solid knowledge of microbiological concepts in the domain of bacterial and plasmid genome analysis. PLACNETw displays the PLACNET graph in a graphical user interface (GUI) and uses simple, comprehensible operations to reconstruct plasmids from WGS datasets. The program is interactive and includes visualization, pruning and saving tools that allow users to easily reconstruct the plasmids in the test genome.

Implementation

For the initial interaction with PLACNETw, the user simply provides pair-end read datasets (two files in gzip fasta format) through the upload page (https://castillo.dicom.unican.es/upload).

PLACNETw is a two-layer web application: server layer and client layer. On the server layer, it uses bash and python to invoke PLACNET and blast queries, thus providing a set of files that will be used at the client layer. On the client layer, it uses ExtJS (https://www.sencha.com/products/extjs) and D3 (https://d3js.org/) for presentation, and a custom-made javascript program for system integration and processing. The interaction between server and client layers uses ajax requests. PLACNETw, which is partly based on a PLACNET Perl script (https://sourceforge.net/projects/placnet/), additionally integrates read assembly and scaffolding detection. Read assembly is performed with Velvet (Zerbino, 2010) through Velvet Optimiser script (https://github.com/tseemann/VelvetOptimiser) to select the optimal kmer length (screening the last 32 bp of maximum read length). Insert sizes and deviations are inferred by Velvet. PLACNETw takes these values to search for potential scaffold links between assembled contigs with Bowtie2 software (Langmead and Salzberg, 2012). Contigs are blasted against a periodically updated reference database, comprising all complete genomes and plasmids from the NCBI database, in order to search for close references. Potential coding sequences are defined with Prodigal software (Hyatt *et al.*, 2010) and then mapped to plasmid protein reference databases for relaxases, replication initiation proteins and incompatibility groups.

The GUI is a three window application (Figure 1). The main window (nº2 in Fig.1) displays the network as a graph, rendered with a force-directed algorithm from D3. In the graph, two types of nodes (assembled contigs and reference genomes) and two types of edges (scaffold links and homology to references) are represented. Contig nodes include a color code, so that plasmid-specific genes (relaxases and replication initiator proteins) stand out in different colors for easier inspection. At the top of the main window there are two button toolbars. The toolbar at the left (nº4 in Fig.1) allows interaction with the graph without modifying it. From left to right, force-directed rendering of the graph can be paused and restarted, contigs can be selected by node length, or reference nodes by string match, relevant information (number and type of selected nodes, and total length of selected contigs) can be obtained about the current selection, the graph can zoom in and out, or rotated in both senses. The toolbar at the right (nº5 in Fig.1) provides tools to modify, save and load the graph. From left to right, a given node can be duplicated to, for example, associate one copy to the main chromosome and the other to one of the plasmids, a new link between two unlinked nodes can be created, the current selected items can be deleted (both nodes and links), previous actions undone, the current state of the network can be saved for future work, the original network can be uploaded, or the results downloaded to the user machine. Additionally, the user can select or unselect both nodes and links in the main window by clicking with the mouse and by drawing a selection area. The window at the left (nº1 in Fig.1) is a node tree linked to the graph in the main window. It represents the network in a hierarchical way. The tree is organized in such a way that contigs are the main nodes and linked references are represented as their children. When the user selects a node from the graph in the main window, it is highlighted in the tree and vice versa. The window at the bottom (nº6 in Fig.1) displays three alternative panels. The first panel shows the result of a BLAST query when the user selects a specific reference in the tree view. The second panel is a notebook that allows the user to take notes for further reference. The third panel is a log of the actions performed by the user (deletions, duplications, etc.) for further reference.

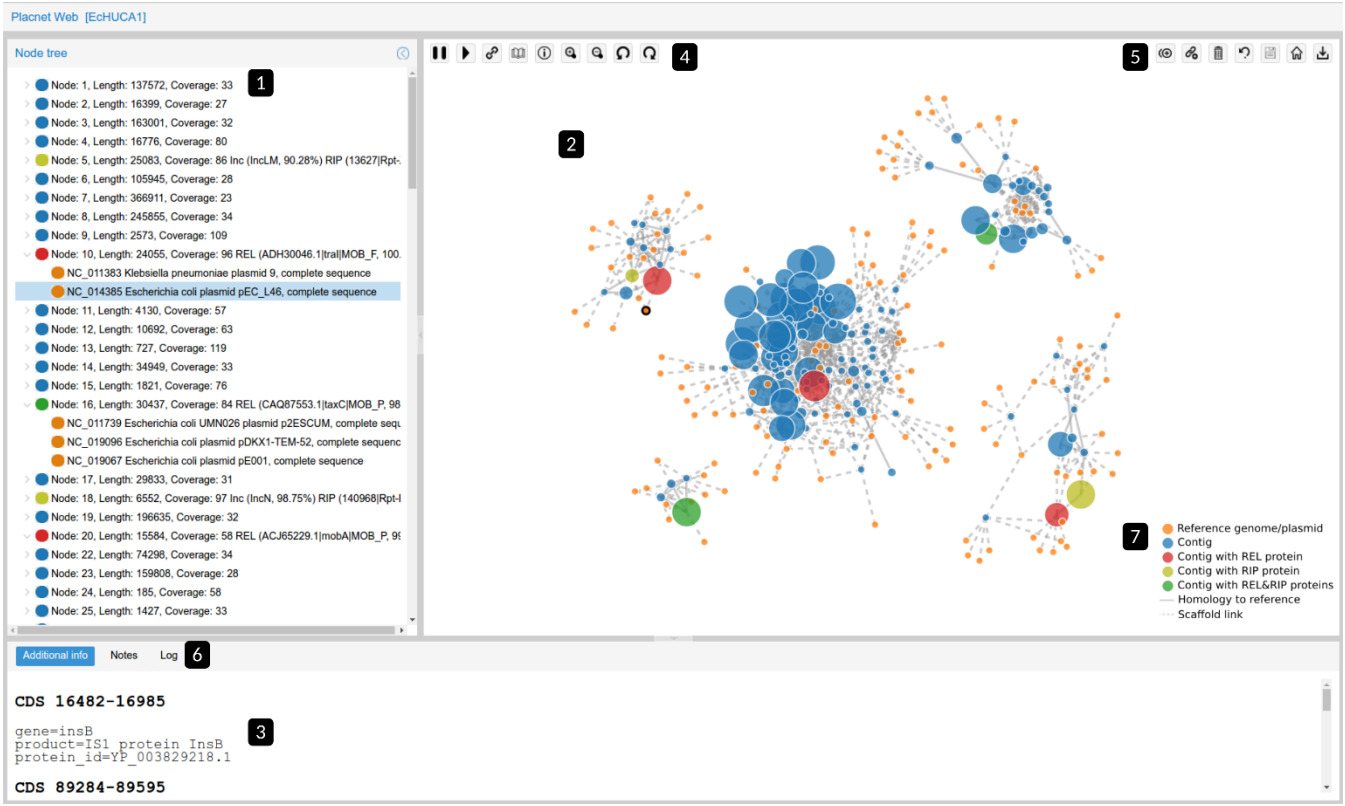

Figure 1. PLACNETw screen of the reconstruction of the *E. coli* L58 genome. L58 is a clinical *E. coli* strain coding for OXA-48 and sequenced by R. Rodicio (University of Oviedo, Spain). PLACNETw plasmid reconstruction results in four large plasmids (120, 62, 41 and 34 kb; data summarized in the L58 example in the PLACNETw web site). The following panels and elements are numbered as shown: (1) Panel displaying a list of contigs and their closest references. (2) Working panel showing a graphic representation of the network. (3) Selected contig or reference information. (4) Visualization tools. (5) Pruning and saving tools. (6) Information panel, composed of three independent tabs: (6a) Additional info: shows Blast results for each contig against its closest references; (6b)

Notes: personal added notes; (6c) Log: record of pruning steps. (7) Network color code.

When the user has finished pruning the network and has identified both chromosome and plasmids, the results can be downloaded from the toolbar as a compressed zip file that contains the following set of files:

- The logfile of PLACNETw run at the server
- A set of files *comp[n].fa*, where n is an integer, with the sequences of each of the connected subgraphs ordered by size (so that n=0 corresponds to the largest component, the chromosome) and other n values correspond to the different plasmids in the genome.
- A text file (notes.txt) with the notes taken by the user inside PLACNETw.
- A text file (log.txt) with the log of user actions inside PLACNETw.
- A text file, named AAAA_Logfile.txt, where AAAA is the date and time where contigs were assembled. This file is produced by Velvet assembler as a logout file. It details contig parameters, such as contig number, N50, longest contig and number of contigs larger than 1 Kb.
- Three tabular files (tempInc.blast, tempRIP.blast, tempREL.blast) showing BLAST results against PLACNETw databases for Incompatibility groups, Replication Initiator Proteins and RELaxases. Further information about BLAST tabular output files can be found at http://www.ncbi.nlm.nih.gov/books/NBK279690/.
- Two text files (placnet.prod.cds, placnet.prod.faa) showing Prodigal (http://prodigal.ornl.gov/) predictions for CDS and proteins from assembled contigs.
- A Scalable Vector Graphics (SVG) file (placnet.svg) which contains an image of the network for publishing or documenting.

Conclusion

PLACNETw is a user-friendly solution for plasmid reconstruction. It builds on different characteristics of plasmid biology to improve the result of contig assembly. This is achieved by manual pruning of the graph representation, which is now optimized in PLACNETw for easiness of use. When all the info is integrated, each plasmid sequence is downloaded as a separate file that can be used in downstream applications.

Funding

This work was supported by the Spanish Ministry of Economy and Competitivity (BFU2011-26608) and the European Seventh Framework Program (612146/FP7-ICT-2013-10).

Conflict of Interest: none declared.

References

Antipov,D. *et al.* (2016) plasmidSPAdes : Assembling Plasmids from Whole. 2014–2015.

Arredondo-Alonso,S. *et al.* (2017) On the (im)possibility to reconstruct plasmids from whole genome short-read sequencing data. *bioRxiv*.

de Been,M. *et al.* (2014) Dissemination of Cephalosporin Resistance Genes between Escherichia coli Strains from Farm Animals and Humans by Specific Plasmid Lineages. *PLoS Genet.*, **10**, e1004776.

Carattoli,A. *et al.* (2014) In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. *Antimicrob. Agents Chemother.*, **58**, 3895–3903.

Hyatt,D. *et al.* (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, **11**, 119.

Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–9.

Lanza,V.F. *et al.* (2014) Plasmid Flux in Escherichia coli ST131 Sublineages, Analyzed by Plasmid Constellation Network (PLACNET), a New Method for Plasmid Reconstruction from Whole Genome Sequences. *PLoS Genet.*, **10**, e1004766.

Leplae,R. *et al.* (2010) ACLAME: a CLAssification of Mobile genetic Elements, update 2010. *Nucleic Acids Res.*, **38**, D57-61.

Treangen,T.J. and Salzberg,S.L. (2011) Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.*, **13**, 36–46.

Wiedenbeck,J. and Cohan,F.M. (2011) Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol. Rev.*, **35**, 957–976.

Zerbino,D.R. (2010) Using the Velvet de novo assembler for short-read sequencing technologies. *Curr. Protoc. Bioinforma.*, **Chapter 11**, Unit 11.5.