BIOINFORMATICS APPLICATIONS NOTE
Vol. 25 no. 5 2009, pages 664–665
doi:10.1093/bioinformatics/btp030

Genome analysis

IslandViewer: an integrated interface for computational identification and visualization of genomic islands

Morgan G. I. Langille and Fiona S. L. Brinkman*

Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada

Received on October 18, 2008; revised on December 12, 2008; accepted on January 12, 2009

Advance Access publication January 16, 2009

Associate Editor: Dmitrij Frishman

ABSTRACT

Summary: Genomic islands (clusters of genes of probable horizontal origin; GIs) play a critical role in medically important adaptations of bacteria. Recently, several computational methods have been developed to predict GIs that utilize either sequence composition bias or comparative genomics approaches. IslandViewer is a web accessible application that provides the first user-friendly interface for obtaining precomputed GI predictions, or predictions from user-inputted sequence, using the most accurate methods for genomic island prediction: IslandPick, IslandPath-DIMOB and SIGI-HMM. The graphical interface allows easy viewing and downloading of island data in multiple formats, at both the chromosome and gene level, for method-specific, or overlapping, GI predictions.

Availability: The IslandViewer web service is available at http://www. pathogenomics.sfu.ca/islandviewer and the source code is freely available under the GNU GPL license.

Contact: brinkman@sfu.ca

1 INTRODUCTION

Large-genomic regions that contain multiple genes of probable horizontal origin, termed genomic islands (GIs), are of significant medical interest because they disproportionately contain genes involved in virulence, antibiotic resistance or other important adaptations (Dobrindt *et al*., 2004; Hacker and Kaper, 2000). Therefore, the identification of GIs has become a particular focus when examining a bacterial genome for its notable new properties. Several computational tools have been developed to predict islands in sequenced genomes (Hsiao *et al*., 2005; Rajan *et al*., 2007; Tu and Ding, 2003; Vernikos and Parkhill, 2006; Waack *et al*., 2006). The majority of these tools utilize the naturally occurring genome sequence biases that exist between bacterial strains to identify regions that appear to have a foreign sequence composition (Karlin, 2001; Vernikos and Parkhill, 2008). In contrast to searching for such anomalous regions using sequence composition signatures, GIs can also be predicted using a comparative genomics approach: identifying regions that have a limited phylogenetic distribution through a comparison of multiple genomes that suggests that the region likely has horizontal origins (Chiapello *et al*., 2005; Langille *et al*., 2008; Ou *et al*., 2006). We now present IslandViewer, the first web accessible interface that facilitates viewing and downloading of GI datasets predicted from user-submitted sequences, or based on precomputed analyses, using sequence composition-based approaches SIGI-HMM and IslandPath-DIMOB, and the comparative genomics approach IslandPick.

*To whom correspondence should be addressed.

2 IMPLEMENTATION

2.1 Genome data source and storage

All sequenced genomes are downloaded from the National Center for Biotechnology Information (NCBI) FTP server (ftp://ftp.ncbi.nih.gov/genomes/Bacteria) each month and loaded into a local MySQL database. GI predictions are precomputed using SIGI-HMM, IslandPath-DIMOB and IslandPick (see below) and are stored so that predictions are available for all new, complete genomes. All methods are run in parallel for each genome so that updates are quickly performed on a computer cluster, while all dynamic web pages are implemented using PHP.

2.2 Genomic island prediction methods

The inclusion of particular GI prediction methods into IslandViewer were based on several factors. The most obvious is that we could only consider including methods that have obtainable software and could be run without manual intervention. Therefore, many GI resources that are simply a database and have no downloadable software such as Islander (Mantri and Williams, 2004) could not be included into IslandViewer. In addition, we did not consider the inclusion of MobilomeFINDER (Ou *et al*., 2007) or MOSIAC (Chiapello *et al*., 2005), two tools that use comparative genomics-based approaches similar to IslandPick because they require the manual selection of comparison genomes (making precomputed results for all genomes impossible). However, all of these methods are listed on the 'Resources' page and we would recommend users visit their respective web sites if interested.

For those tools that did have their software freely available, we included IslandPath-DIMOB (Hsiao *et al*., 2005) and SIGI-HMM (Waack *et al*., 2006) because they were shown to have the highest specificity (86–92%) and overall accuracy (86%) (Langille *et al*., 2008). In addition, we included the automated comparative genomics method, IslandPick, since it provides predictions that are not based on sequence composition and showed the most agreement with a manual curated dataset of literature-based GIs (Langille *et al*., 2008). These three methods sometimes predict the same GIs, but often give slightly different results suggesting that they complement each other well without being redundant. We avoided

© 2009 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

the inclusion of other methods that had lower specificity (some as low as 38% precision), which would result in a large number of false predictions in IslandViewer. Finally, none of the methods included in IslandViewer had been previously available as a web resource; therefore, giving new user-friendly access to three different GI prediction methods.

2.3 IslandViewer interface

IslandViewer allows the viewing of all GI predictions for the above predictors through a single integrated interface. Predictions are precomputed for all published GIs and are updated on a monthly basis, while users with newly sequenced unpublished genomes can submit their genome for analysis and receive an email notification when finished. These user-submitted genomes are not viewable by other IslandViewer users and are accessible for at least 1 month. IslandPick automatically selects comparison genomes for use using default distance parameters, but since researchers may have particular insights into a particular species, they can choose to run IslandPick with their own manually selected comparison genomes and have the option of being notified by email when the results are available.

Once the genome of interest is selected it is presented as a circular genome image with each predicted GI highlighted (different colours for different tools in the IslandViewer) and is also available as a high-resolution image suitable for publication. In addition to the predicted GIs for each tool, IslandViewer highlights any GIs that have been predicted by two or more methods. The annotations for genes within each GI can be quickly viewed by hovering over the GI of interest within the image. Clicking on an island jumps to the corresponding row in a table below the genome image and gives information such as GI coordinates, links to tables showing genes and annotations within the GI region, links to external genome viewers at NCBI and joint genome institute (JGI), and links to IslandPath to allow further examination of GI-related features in the genome of choice. GI predictions may be downloaded in various formats including Excel, tab-delimited, comma-delimited, Fasta and Genbank (allowing easy input into the genome browser and annotation tool Artemis). In addition, we provide a 'Resources' page that links to other GI prediction methods that are not included in IslandViewer, but may be useful to users who wish to investigate different prediction methods. All datasets and source code are available for download under a GNU GPL license.

3 CONCLUDING COMMENTS

GI identification is becoming a first critical step in the characterization of a bacterial genome, due to the growing appreciation for the role of GIs in important adaptations of interest. Recent research has therefore focused on developing new computational methods for their prediction. However, these methods tend to use different approaches and identify different features of GIs. The result is that the most accurate methods each have high precision, but low recall, leading to slightly different regions being predicted. Previously, researchers could either pick a single method or try to manually integrate the results from multiple methods themselves. In addition, many of these tools did not have their own web interfaces and often required that the user download and run the program on their computer. IslandViewer alleviates these concerns by providing a web interface for three accurate GI prediction methods that were not previously available through a web interface. By precomputing GI datasets for all completed genomes and providing a single submission process for new user genomes, we allow researchers access to a user-friendly resource that can be used as the first step in GI analysis of bacterial genomes. We would expect that researchers would manually inspect any GI predictions shown in IslandViewer to determine their validity and make more accurate predictions of their boundaries. IslandViewer helps aid further analysis of GI predictions by providing data in various formats that can be used in other bioinformatic tools such as Artemis, and by providing numerous links to other GI resources. IslandViewer should be a useful resource for any researcher studying GIs and microbial genomes.

ACKNOWLEDGEMENTS

M.G.I.L. also holds a MSFHR scholarship, while F.S.L.B. is a MSFHR Senior Scholar and CIHR New Investigator. Infrastructure support was also provided by Genome Canada/GenomeBC, SFU CTEF and IBM.

Funding: Canadian Institutes of Health Research and Michael Smith Foundation for Health Research (for SFU/UBC Bioinformatics Training Program).

Conflict of Interest: none declared.

REFERENCES

Chiapello,H. *et al.* (2005) Systematic determination of the mosaic structure of bacterial genomes: species backbone versus strain-specific loops. *BMC Bioinformatics*, **6**, 171.

Dobrindt,U. *et al.* (2004) Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.*, **2**, 414–424.

Hacker,J. and Kaper,J.B. (2000) Pathogenicity islands and the evolution of microbes. *Ann. Rev. Microbiol.*, **54**, 641–679.

Hsiao,W.W. *et al*. (2005) Evidence of a large novel gene pool associated with prokaryotic genomic islands. *PLoS Genet.*, **1**, e62.

Karlin,S. (2001) Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes. *Trends Microbiol.*, **9**, 335–343.

Langille,M.G.I. *et al*. (2008) Evaluation of genomic island predictors using a comparative genomics approach. *BMC Bioinformatics*, **9**, 329.

Mantri,Y. and Williams,K.P. (2004) Islander: a database of integrative islands in prokaryotic genomes, the associated integrases and their DNA site specificities. *Nucleic Acids Res.*, **32**, D55–D58.

Ou,H.Y. *et al.* (2006) A novel strategy for the identification of genomic islands by comparative analysis of the contents and contexts of tRNA sites in closely related bacteria. *Nucleic Acids Res.*, **34**, e3.

Ou,H.Y. *et al.* (2007) MobilomeFINDER: web-based tools for in silico and experimental discovery of bacterial genomic islands. *Nucleic Acids Res.*, **35**, W97–W104.

Rajan,I. *et al*. (2007) Identification of compositionally distinct regions in genomes using the centroid method. *Bioinformatics*, **23**, 2672–2677.

Tu,Q. and Ding,D. (2003) Detecting pathogenicity islands and anomalous gene clusters by iterative discriminant analysis. *FEMS Microbiol. Lett.*, **221**, 269–275.

Vernikos,G.S. and Parkhill,J. (2006) Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the Salmonella pathogenicity islands. *Bioinformatics*, **22**, 2196–2203.

Vernikos,G.S. and Parkhill,J. (2008) Resolving the structural features of genomic islands: a machine learning approach. *Genome Res.*, **18**, 331–342.

Waack,S. *et al*. (2006) Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics*, **7**, 142.