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

Genome analysis

MeltingPlot: a user-friendly online tool for epidemiological investigation using High Resolution Melting data

Matteo Perini[1], Gherard Batisti Biffignandi[2], Domenico Di Carlo[1], Ajay Ratan Pasala[1], Aurora Piazza[2], Simona Panelli[1], Gian Vincenzo Zuccotti[1,3], and Francesco Comandatore[1,*]

[1]Department of Biomedical and Clinical Sciences "L. Sacco", University of Milan, Pediatric Clinical Research Center "Romeo and Enrica Invernizzi", Milan, 20157, Italy; [2]Department of Clinical, Surgical, Diagnostic and Pediatric Sciences, University of Pavia, Pavia, 27100, Italia; [3]Department of Pediatrics, Children's Hospital Vittore Buzzi, University of Milan, Italy;

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Summary: MeltingPlot is an open source web tool for pathogen typing and epidemiological investigations using High Resolution Melting (HRM) data. The tool implements a graph-based algorithm designed to discriminate pathogen clones on the basis of HRM data, producing portable typing results. MeltingPlot also merges typing information with isolates and patients metadata to create graphical and tabular outputs useful in epidemiological studies. HRM technique allows pathogen typing in less than 5 hours with ~5 euros per sample. MeltingPlot is the first tool specifically designed for HRM-based epidemiological studies and it can analyse hundreds of isolates in a few seconds. Thus, the use of MeltingPlot makes HRM-based typing suitable for large surveillance programs as well as for rapid outbreak reconstructions.

Availability and implementation: MeltingPlot is implemented in R.
The web interface is available at https://skynet.unimi.it/index.php/tools/meltingplot.
The source code is also available at https://github.com/MatteoPS/MeltingPlot.

Contact: francesco.comandatore@unimi.it

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

The rapid typing of pathogens is pivotal to perform fast epidemiological investigations and to detect outbreaks. High-Resolution Melting (HRM) analysis is a molecular biology technique suitable for the aim (Ruskova and Raclavsky, 2011; Nakagawa *et al.*, 2009), allowing pathogen typing in less than 5 hours. For each isolate, HRM analyses *n* specific genomic regions returning *n* melting temperatures (each genomic region is defined by a specific PCR primer set). These temperatures can be used to cluster the isolates in a *n*-dimensional space. Previously, we developed a graph-based algorithm for isolate clustering on the basis of HRM temperatures, and we successfully tested this approach on 82 isolates of *Klebsiella pneumoniae* (Perini et al., 2020), one of the most important nosocomial pathogen world-wide (David *et al.*, 2019). A limit of this HRM clustering approach is that the obtained clusters are valid only for the analysed dataset and thus the output is not portable among laboratories.

Here we present MeltingPlot, a tool for rapid epidemiological investigation using HRM data. The tool implements an evolution of the clustering algorithm we already published (Perini et al., 2020) and it allows to obtain a portable HRM-based typing. Furthermore, the tool merges HRM typing information with isolates/patients metadata to get a complete epidemiological investigation. MeltingPlot has a user-friendly web interface (the standalone version is also available) and it creates easy to read graphical and tabular outputs. The tool runs in a few seconds even with hundreds of isolates.

2 Description

The MeltingPlot flow can be divided in three main steps**:** HRM-based clustering/typing of isolates, prevalence analysis and transmission analysis. The first step requires HRM temperatures only, while the other two require isolates and patients metadata.

Input preparation. Users are required to download an xls template spreadsheet and fill it with data. The spreadsheet contains three sheets: *HRM_temperatures*, *Isolates_metadata* and *Reference_isolates*.

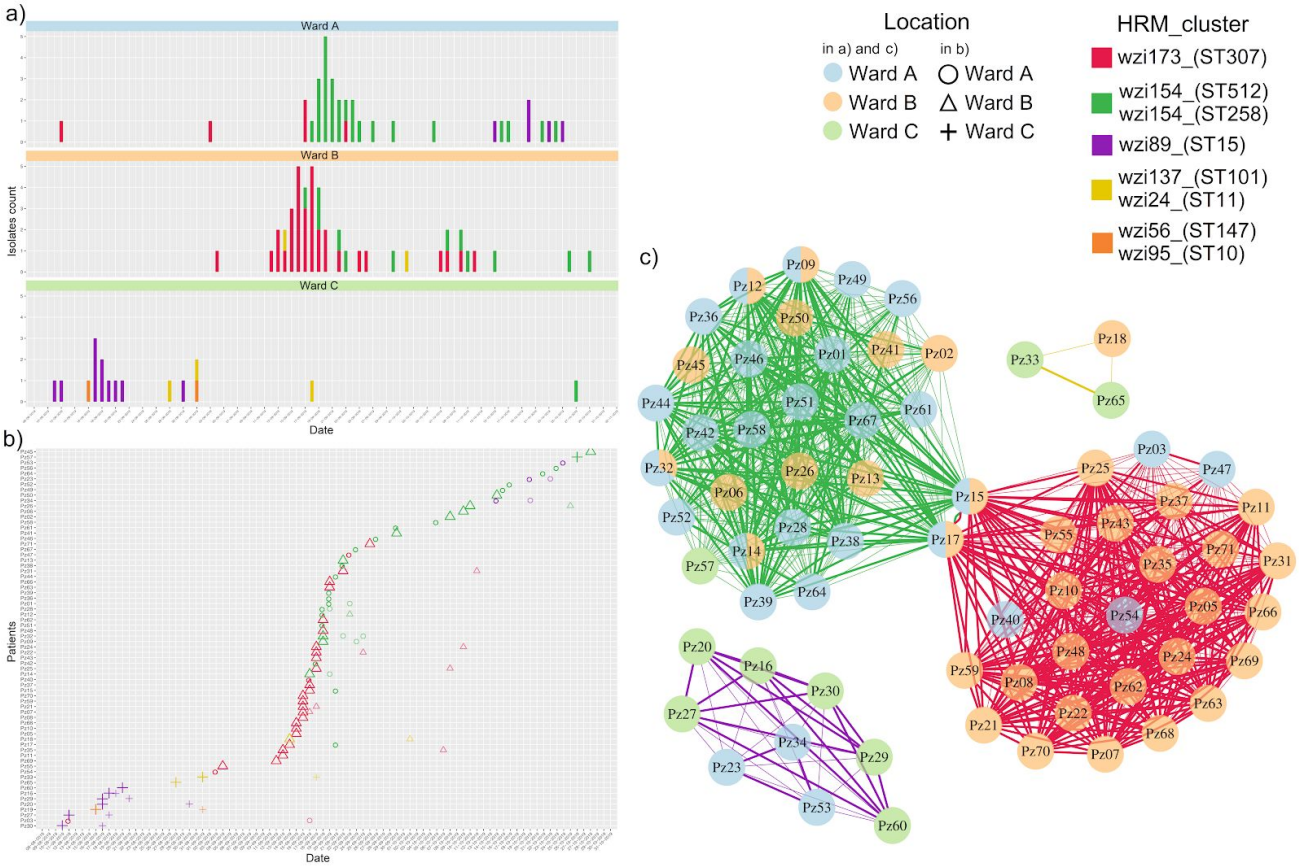

Fig. 1. MeltingPlot Output, example of an epidemiological investigation. Here we report three plots from the MeltingPlot output, selected to show the power of the tool. **a)** Prevalence analysis: the plot shows the number of isolates collected from each hospital ward over time. Each HRM cluster is represented with a different color. This analysis allows the real time detection of the pathogen clones emergence in the hospital setting. **b)** Patients' timeline: each row is a patient and symbols refer to isolates. The symbols shape represent the location where the isolates were collected while the colors represent the isolates' HRM clusters. **c)** Patient-to-patient graph: vertices represent patients and they are reported as pie charts to show the locations (wards) where patients' isolates were collected. Two vertices are connected by an edge if isolates from the same HRM cluster were collected from both the patients. Edges are thicker if the isolates from the same HRM cluster were collected within seven days (this threshold can be defined by the user) in the same location. This plot can help to identify the transmission routes of the pathogen in the hospital setting.

HRM_temperatures: in this sheet the users have to report the HRM temperatures that they obtained from their HRM experiments. If HRM experiments were performed using technical replicates, the users have to provide each replicate temperature (this is the only mandatory data).

Isolates_metadata: in this sheet the users can provide information on the isolates, i.e. isolation date, isolation location (e.g. hospital ward) and an ID for the patients (e.g. Pz.1,...). This information is not mandatory for HRM isolates typing but is required to perform the complete epidemiological investigation.

Reference_isolates: this sheet contains the HRM temperatures of the reference isolates and their annotation (e.g. WGS typing information). The use of the same reference collection in different laboratories allows to obtain portable HRM typing results. MeltingPlot templates with reference HRM temperature collections are available on the tool web site.

HRM-based Clustering/Typing. Isolates clustering is computed on the basis of the melting temperatures. The isolates are organized in a graph where the vertices are the isolates and two vertices are connected if the isolates HRM temperatures differ less than 0.5 °C for each PCR primer set. The graph is then decomposed into separate components and clusters are detected by the Girvan-Newman algorithm (Newman and Girvan, 2004). The isolates connected by edges with a normalized betweenness value above a threshold (this parameter can be set by the user, default is 0.5) will be not assigned to any cluster and classified as "undetermined". If reference HRM data is provided, the clusters that contain one or more reference isolates will be named with the reference isolates annotations (e.g. Sequence Type). For details see Supplementary Material.

Outputs. MeltingPlot creates three groups of plots: the HRM-based clustering/typing, the prevalence analysis and the transmission analysis. The first plot group includes the isolates graph (with nodes colored on the basis of the clusters) and a heatmap showing HRM temperatures and the isolates clusters. These last two groups are created if isolates metadata are provided. The prevalence analysis group includes bar plots showing the distribution of the clusters over time in the different locations. The transmission analysis group contains a patient timeline and a patient-to-patient graph. See Fig. 1 and Supplementary Material for details. MeltingPlot also provides the xls spreadsheets containing isolates HRM clusters and metadata. (see Supplementary Material for detail).

Example of epidemiological investigation. We simulated a large *K. pneumoniae* outbreak (100 isolates) sustained by multiple clones, a situation observed in real nosocomial outbreaks (Ferrari *et al.*, 2019). HRM temperatures were extracted from a dataset of *K. pneumoniae* isolates previously analyzed in our laboratory. We used as reference collection 18 representative isolates out of the 82 typed by HRM and

WGS in Perini *et al.*, 2020 (this collection is available on the tool web site). In a real situation, the HRM typing of the 100 isolates would be performed day by day for the entire outbreak period (~ 3 months) for a total cost of ~500 euros. As shown in Fig. 1, the outbreak is sustained by three major isolate clusters. Using the annotation of the reference isolates MeltingPlot labelled these cluster as wzi173_(ST307) (in red), wzi154_(ST512/ST258) (in green) and wzi89(ST15) (in violet), giving a portable typing. The main outbreak is caused by the two pathogen clusters coloured in green and red. Each of these clusters is highly associated with a single ward, the green one with Ward A and the red one with Ward B. The violet cluster causes a small outbreak in Ward C. The patient's timeline and the patient-to-patient graph clearly show that two patients (Pz15 and Pz17) infected by isolates of the red and green clusters cross the wards A and B, showing a possible transmission route among the wards. MeltingPlot provides a general view of the pathogen spreading in the hospital setting allowing the reconstruction of possible transmission routes. A complete description of each output is available in the Supplementary Material.

Acknowledgements

Thanks to the Romeo ed Enrica Invernizzi Foundation.

Funding

Conflict of Interest: none declared.

References

David,S. *et al.* (2019) Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial spread. *Nat Microbiol*, **4**, 1919–1929.

Ferrari,C. *et al.* (2019) Multiple KPC Clones Contribute to an Extended Hospital Outbreak. *Front. Microbiol.*, **10**, 2767.

Nakagawa,N. *et al.* (2009) Cocirculation of antigenic variants and the vaccine-type virus during the 2004-2005 influenza B virus epidemics in Japan. *J. Clin. Microbiol.*, **47**, 352–357.

Newman,M.E.J. and Girvan,M. (2004) Finding and evaluating community structure in networks. *Physical Review E*, **69**.

Perini,M. *et al.* (2020) EasyPrimer: user-friendly tool for pan-PCR/HRM primers design. Development of an HRM protocol on *wzi* gene for fast Klebsiella pneumoniae typing. *Sci. Rep.*, **10**, 1307.

Ruskova,L. and Raclavsky,V. (2011) The potential of High Resolution Melting Analysis (HRMA) to streamline, facilitate and enrich routine diagnostics in medical microbiology. *Biomedical Papers*, **155**, 239–252.