Bayesian prediction of bacterial growth temperature range based on genome sequences

Background: The preferred habitat of a given bacterium can provide a hint of which types of enzymes of potential industrial interest it might produce. These might include enzymes that are stable and active at very high or very low temperatures. Being able to accurately predict this based on a genomic sequence, would thus allow for an efficient and targeted search for production organisms, reducing the need for culturing experiments. Results: This study found a total of 40 protein families useful for distinction between three thermophilicity classes (thermophiles, mesophiles and psychrophiles). The predictive performance of these protein families were compared to those of 87 basic sequence features (relative use of amino acids and codons, genomic and 16S rDNA AT content and genome size). When using naive Bayesian inference, it was possible to correctly predict the optimal temperature range with a Matthews correlation coefficient of up to 0.68. The best predictive performance was always achieved by including protein families as well as structural features, compared to either of these alone. A dedicated computer program was created to perform these predictions. Conclusions: This study shows that protein families associated with specific thermophilicity classes can provide effective input data for thermophilicity prediction, and that the naive Bayesian approach is effective for such a task. The program created for this study is able to efficiently distinguish between thermophilic, mesophilic and psychrophilic adapted bacterial genomes.

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, Novozymes AS
Contributors: Jensen, D. B., Vesth, T. C., Hallin, P. F., Pedersen, A. G., Ussery, D.
Publication date: 2012
Peer-reviewed: Yes

Publication information
Journal: B M C Genomics
Volume: 13
Issue number: Suppl. 7
ISSN (Print): 1471-2164
Ratings:
BFI (2019): BFI-level 1
Web of Science (2019): Indexed yes
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): CiteScore 4.08 SJR 2.11 SNIP 1.151
Web of Science (2017): Impact factor 3.73
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 4.05 SJR 2.163 SNIP 1.096
Web of Science (2016): Impact factor 3.729
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): CiteScore 4.3 SJR 2.348 SNIP 1.159
Web of Science (2015): Impact factor 3.867
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): CiteScore 4.18 SJR 2.327 SNIP 1.199
Web of Science (2014): Impact factor 3.986
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): CiteScore 4.39 SJR 2.195 SNIP 1.188
Web of Science (2013): Impact factor 4.041
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): CiteScore 4.61 SJR 2.236 SNIP 1.243
Web of Science (2012): Impact factor 4.397