Amino Acids (2014) 46:471–473
DOI 10.1007/s00726-013-1649-z

SHORT COMMUNICATION

# SECRETOOL: integrated secretome analysis tool for fungi

**Ana R. Cortázar · Ana M. Aransay · Manuel Alfaro · José A. Oguiza · José L. Lavín**

Received: 8 November 2013 / Accepted: 12 December 2013 / Published online: 27 December 2013


**Abstract** The secretome (full set of secreted proteins) has been studied in multiple fungal genomes to elucidate the potential role of those protein collections involved in a number of metabolic processes from host infection to wood degradation. Being aminoacid composition a key factor to recognize secretory proteins, SECRETOOL comprises a group of web tools that enable secretome predictions out of aminoacid sequence files, up to complete fungal proteomes, in one step. SECRETOOL is freely available on the web at http://genomics.cicbiogune.es/SECRETOOL/Secretool.php.

**Keywords** Protein secretion · Signal peptide · Secretome · Bioinformatics · Analysis pipeline

## Introduction

Most eukaryotic proteins are synthesized in the cytosol and eventually may be sorted to different subcellular locations. Extracellular proteins contain an N-terminal sequence that is recognized by the secretory pathway. These signal peptides (SPs) mediate targeting proteins to the endoplasmic reticulum for subsequent transport through the secretory pathway (Rapoport 1992). In most of the secreted proteins, SPs are cleaved off by specific signal peptidases. In order to reach the outer surface of the fungus, secreted or extracellular proteins have to travel through the cell wall, and, therefore, can be found in the in vitro growth medium (Fonzi 2009). Next to the growth medium, the cell wall is also regarded as an extracellular entity (Sorgo et al. 2013). To perform an in silico classical secretome analysis (Wymelenberg et al. 2005; Brown et al. 2012), SECRETOOL allows the screening of different features for each of the proteins annotated as part of the fungus proteome (Lum and Min 2011): (1) location of the predicted N-terminal SP, (2) detection of the presence and location of SP cleavage sites in amino acid sequences, (3) presence of a maximum of one transmembrane domain (TMD) and (4) GPI membrane anchoring. In addition, a non-classical analysis can be carried out with the presented tool by searching for not signal peptide-triggered protein secretion, trying to offer a wider range of predictions. It has to be highlighted that fungal genome annotation in databases needs still to be refined and that a substantial percentage of the predicted models might be truncated at 5′ termini with no ATG present. Nevertheless, the web-based SECRETOOL is able to predict putative secreted proteins and infer their domain structure and ortholog relations among fungi.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00726-013-1649-z) contains supplementary material, which is available to authorized users.

A. R. Cortázar · A. M. Aransay · J. L. Lavín (✉)
Genome Analysis Platform, CIC bioGUNE and CIBERehd, Bizkaia Technology Park, 48160 Derio, Spain
e-mail: jllavin@cicbiogune.es

M. Alfaro · J. A. Oguiza
Department of Agrarian Production, Genetics and Microbiology Research Group, Public University of Navarre, 31006 Pamplona, Spain

## Materials and methods

### Implementation

SECRETOOL has been placed under a Linux environment in an Apache server, which provides a stable and secure background. A Perl/CGI control module has been

developed to enable communication between the user interface based on PHP/CSS and the tools within the server. Prediction tools included in SECRETOOL are based only upon publicly available software: *TargetP*, *SignalP* (Emanuelsson et al. 2007), *PredGPI* (Pierleoni et al. 2008), *TMHMM* (Krogh et al. 2001), *WoLFPSORT* (Horton et al. 2007), *SecretomeP* (Bendtsen et al. 2004), *BLAST+* (Camacho et al. 2009) and *HMMER3* (Finn et al. 2011). A set of Perl/Bioperl scripts has been designed to carry out complementary tasks and to connect the individual steps of pipelines for proteome scale analyses.

## Results and discussion

SECRETOOL is supported by different types of custom local tools, but also offers a pool of tools for complementary remote analyses (Supplementary Table 1). These tools have been tested on 70 different fungal proteomes (Supplementary Table 2) in order to calibrate the best performance with this particular kind of input. At some extent other eukaryotic proteomes (mainly animal) may be screened by this tool with an acceptable performance, but prokaryotic genomes are not covered by SECRETOOL.

### Local tools

The local tools of SECRETOOL are oriented to one-step analyses and can be classified into two categories: full analysis pipeline and independent analysis tools.

#### *Full analysis pipeline*

In order to perform a complete secretome analysis in an intuitive manner, we have set up a pipeline including the most accepted tools for secretome analysis in fungi along with some customized Perl/Bioperl scripts (Fig. 1). Tools for classical secretion analysis are presented under a one-step query interface, where users upload an input file in FASTA format (as big as a full proteome), select the cutoff scores for the processing tools (or default values) and submit data. This pipeline begins processing the data with *TargetP, SignalP* and *PredGPI*. Then, proteins predicted by these three methods are merged into a common list (in order not to lose false negative predictions that could be filtered if they were just intersected) to be considered downstream for TMD evaluation by *TMHMM*. Candidate proteins with 0 or 1 TMD are kept as input for *WoLFPSORT*, where sequences labeled as extracellular are retained. These prediction steps yield a list with the Ids of the putatively secreted proteins and a file where the sequences relative to those Ids are stored in FASTA format. Finally, the pipeline can carry out the orthology prediction

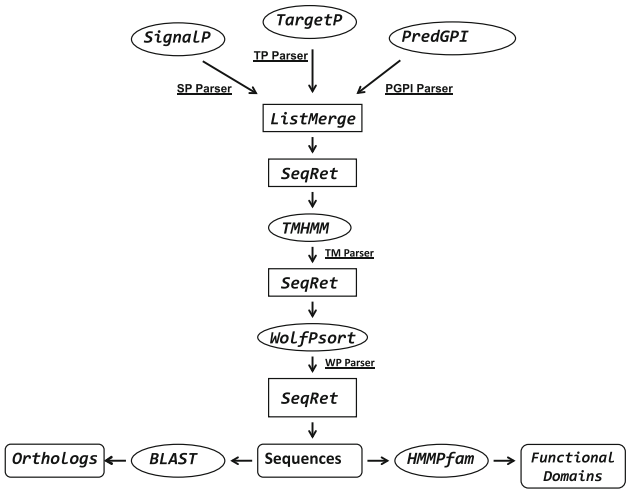


**Fig. 1** Full SECRETOOL workflow. *Ovals* represent pre-installed prediction tools, *rectangles* represent custom tools and result parsers appear underlined. *Rounded rectangles* represent the final output

and the determination of the domain structure. These two steps are offered as optional, since they are the most time-consuming processes of the whole procedure (Supplementary Table 3). If chosen by user, retained sequences are queried into *BLAST+* versus non-redundant database, and into *HMMER3* against *Pfam* database (Finn et al. 2006), respectively. Therefore, the results of running the classical secretion whole pipeline are output in four plain text files: (1) predicted putative secreted proteins Id list, (2) putative secreted protein sequences in FASTA format, (3) tabular BLAST output displaying putative functions and orthology relationships with other fungi and (4) tabular *HMMER3* output showing the domain structure of the protein sequences. Since this process is time-consuming when FASTA files are very long (over 1,000 protein sequences) (Table 1), a link to the results is sent via e-mail to the user upon job completion.

#### *Independent analysis tools*

SECRETOOL includes a set of tools to carry out diverse sequence manipulation tasks and predictions, namely sequence retriever (SEQRET) which retrieves a given list of sequences from a FASTA file of bigger size such as full proteome and *PredGPI* parser that filters predictions on GPI anchor domains from *PredGPI*. In addition, each of the prediction tools used in the full pipeline can be implemented individually; although, in this mode, they are only aimed to identify secreted proteins (other options are disabled). These tools test different parameters (e.g. presence of N-terminal presequences in *TargetP*, SP cleavage sites with *SignalP*, multi-factorial subcellular location predictions by *WoLFPSORT* and the presence of TMDs via

*TMHMM*) for a broader characterisation. *BLAST* and *HMMER3* are also included as separate tools for functional characterization (as in the pipeline). Independent tools seek to maintain the application philosophy of full proteome analysis in one step admitting user-defined cutoff values for improved usefulness.

Alternatively there is an option to run a non-classical secretion analysis based on *SecretomeP* that produces ab initio predictions of non-classical protein secretion (no signal peptide triggered) through a Neural Network to obtain the final secretion prediction. These prediction steps yield a list with the Ids of non-classically secreted proteins and a FASTA file with those sequences.

Remote tools

A compendium of links to each tool web server is available via SECRETOOL through "Remote tools" menu.

## Conclusion

In this work, we describe the web-based analysis pipeline SECRETOOL, which is specifically developed and tested for the in silico screening and identification of putative secreted proteins in fungi (although it may analyze other non-plant eukaryotic genomes). SECRETOOL is a compendium of analysis tools that improves the handling presented by each considered tool independently and as part of a full analysis pipeline. This implementation to perform analysis on user provided data sets was done under a common philosophy: whole pipeline requires just one file submission to run and avoids onerous data preprocessing steps, thus, the need for previous command line experience and sequence number limitations of the precursor tools are overcome. The advantage of using SECRETOOL versus other more database oriented tools (Caccia et al. 2013), for instance, *FunSecKB* (Lum and Min 2011) developed to grant access to pre-computed data, is that this new pipeline is designed to enhance secreted proteins prediction in addition to allow further analyses on the results.

**Acknowledgments** We thank J. Ferrero for his advice with Perl/CGI scripting. ARC, AMA, JLL and the research expenses are supported by the Gov. of the Basque Country (Etortek Research Programs 2011/2013) and from the Innovation Technology Dept. of Bizkaia. MA and JAO are supported by research projects AGL2011-30495 of the Spanish National Research Plan, Bioethanol-Euroinnova from the Autonomous Gov. of Navarre and by additional support from the Public University of Navarre.

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

Bendtsen JD, Jensen LJ, Blom N, von Heijne G, Brunak S (2004) Feature based prediction of non-classical and leaderless protein secretion. Protein Eng Des Sel 17(4):349–356

Brown NA, Antoniw J, Hammond-Kosack KE (2012) The predicted secretome of the plant pathogenic fungus Fusarium graminearum: a refined comparative analysis. PLoS One 7(4):e33731. doi:10.1371/journal.pone.0033731

Caccia D, Dugo M, Callari M, Bongarzone I (2013) Bioinformatics tools for secretome analysis. Biochim Biophys Acta 1834(11):2442–2453

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinformatics 10:421. doi:10.1186/1471-2105-10-421

Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2(4):953–971

Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A (2006) Pfam: clans, web tools and services. Nucleic Acids Res 34(Database issue):D247-251

Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. Nucleic Acids Res 39(Web Server issue):W29-37

Fonzi WA (2009) The protein secretory pathway of *Candida albicans*. Mycoses 52(4):291–303

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K (2007) WoLF PSORT: protein localization predictor. Nucleic Acids Res 35(Web Server issue):W585-587

Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305(3):567–580

Lum G, Min XJ (2011) FunSecKB: the Fungal Secretome KnowledgeBase. Database: bar001. doi:10.1093/database/bar001

Pierleoni A, Martelli PL, Casadio R (2008) PredGPI: a GPI-anchor predictor. BMC Bioinformatics 9:392 doi:10.1186/1471-2105-9-392

Rapoport TA (1992) Transport of proteins across the endoplasmic reticulum membrane. Science 258(5084):931-936

Sorgo AG, Heilmann CJ, Brul S, de Koster CG, Klis FM (2013) Beyond the wall: *Candida albicans* secret(e)s to survive. FEMS Microbiol Lett 338(1):10–17

Wymelenberg AV, Sabat G, Martinez D, Rajangam AS, Teeri TT, Gaskell J, Kersten PJ, Cullen D (2005) The Phanerochaete chrysosporium secretome: database predictions and initial mass spectrometry peptide identifications in cellulose-grown medium. J Biotechnol 118(1):17–34