

SIPHT Characterization

SIPHT

The bioinformatics project at Harvard University is conducting a wide search for small untranslated RNAs (sRNAs) that regulate several processes such as secretion or virulence in bacteria. The sRNA identification protocol using high-throughput technology (SIPHT) program uses a workflow to automate the search for sRNA encoding-genes for all of the bacterial replicons in the National Center for Biotechnology Information (NCBI) database. The kingdom-wide prediction and annotation of sRNA encoding genes involves a variety of individual programs that are executed in the proper order using Condor DAGMan's capabilities. These involve the prediction of Rho-independent transcriptional terminators, BLAST (Basic Local Alignment Search Tools) comparisons of the inter genetic regions of different replicons and the annotations of any sRNAs that are found.

Execution Profile

Execution times of SIPHT jobs			
Job	Count	Mean (s)	Variance
Blast_candidate	1	5.8	0
Blast_paralogues	1	4.5	0
Blast_QRNA	1	1344.88	0
Blast_synteny	1	33	0
FFN_parse	1	1.4	0
Findterm	1	975.16	0
Patser	17	1.3	0.19
Patser_concate	1	0.01	0
RNAMotif	1	44	0
SRNA	1	306.53	0
SRNA_annotate	1	1.9	0
Transterm	1	32	0

Sizes of SIPHT data items			
File Type	Count	Mean (MB)	Variance
alphabet	1	1.5e-05	0
blasta	1	0.75	0
BLAST_*	1	0.0081	0
BLAST_Out	1	2	0
blast_paralogues.out	1	0.003	0
BLAST_sorted.out	1	4.3	0
blast_synteny.out	1	0.011	0
eqrna	1	0.91	0
findterm.out	1	0.00013	0
IGR_Partners	934	0.28	0.098
TFBS_matrices	18	0.0097	0.0014
• .ffn	1	3.9	0
• .fna	1	4.3	0
• .gbk	1	10	0
• _paralogues.txt	1	1.2	0
• _parsed.ffn	1	0.83	0

• _PatserOut.txt	1	0.11	0
• .ptt	1	0.31	0
• _QRNA.txt	1	0.75	0
• _sRNA.out	1	0.098	0
• _sRNA.out_annotated	1	0.2	0
• _synteny.txt	1	1.3	0
• _term_cand_nonredund	1	0.37	0
• _term.txt	1	23	0
OutBlastParsed	1	0.48	0
OutCandidates	3	0.26	0.0076
patser.in	1	1.3e-05	0
patser.out	17	0.0067	2e-05
QRNA_out	1	0.0093	0
RNAMofficial_desc.txt	1	0.017	0
mmotif.out	1	1	0
rna.ps	1	0.0027	0
Seq_known_sRNAs_IGRs	1	0.16	0
Seq_*	1	0.077	0
srna_annotate.out	1	0.65	0
sRNAPredict.in	1	0.0011	0
transterm.err	1	9.2e-05	0
transterm.out	1	0.41	0
vienna_index_tmp	1	6.8	0
vienna_input_tmp	1	4.7	0
vienna_output	1	11	0
xdfformat	1	0.36	0