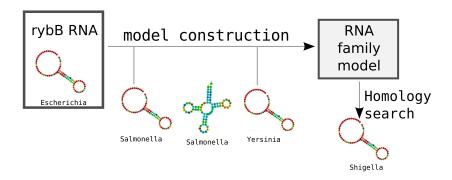
Unsupervised RNA family model building

Florian Eggenhofer

Institute for Theoretical Chemistry University of Vienna

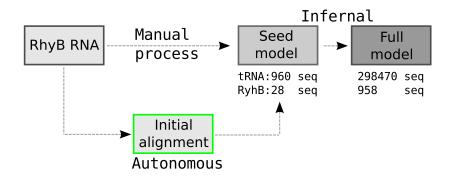
February 19, 2015

RNA family model construction



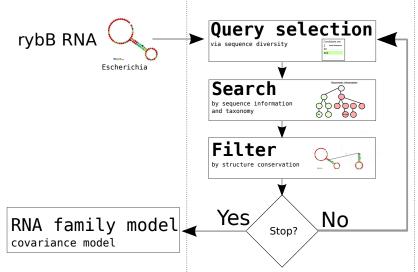
F. Eggenhofer TBI

Challenges in RNA family model construction



F. Eggenhofer TBI

Pipeline - RNAlien



Unsupervised RNA family model building

F. Eggenhofer

Tool

- ► Commandline tool
- ▶ Input: Fasta-File, optional Taxonomy Id
- ▶ Dependencies: infernal, locarna, RNAz, clustal-o
- ▶ Database interfaces via REST
- ► available via Hackage: cabal install RNAlien (soonish)

Webservice - Input



Welcome Earthling

Upload a fasta sequence file						
Browse	No file selected.					
Enter Taxonomy Id:						
Send it! ①	l					

TBI - 2015

Webservice - Output

RNÅlien

Progress:

Iteration	Upper taxonomic treshold id	Inclusion threshold[bit]	Aligned Sequences	Alignment Link	Covariance Model Link	Status
0	279010	not set	0	stockholm-format	covariance-model	done
1	1386	90.07	13	stockholm-format	covariance-model	done
2	186817	90.07	17	stockholm-format	covariance-model	done
3	1385	90.07	24	stockholm-format	covariance-model	done

Results:

Result Fasta Result Alignment Result CM



TBI - 2015

Output

gi[738892387:48578-48811		ACGUUGAGUGAGGCACUAUCCGUC.U.GU.AGACcUGA.UUGUAUUUGUACACCUAUC.UUUGuuuuuaaGGUAGGUGUUUUU
#=GR gi 738892387:48578-48811	PP	***************************************
gi 253987483:c4795122-4794906		ACGUUGAGUGAGCACUAUCC GUC. U. GU. AGAC cUGA. U U G U AUUUAUACACCUAUC. UUUC
#=GR gi 253987483:c4795122-4794906	PP	*********99885*** * * * * * * * * * * * * * * * *
gi 50118965:c4684968-4684740		ACGUUGAGUGAGGCACCAUUC.UGUUGUC.U.GACAGACCUG.AUAUUUUC-AACGCUUACC.GUUUaUCGGUAAGCGUUUUUUU
#=GR gi 50118965:c4684968-4684740	PP	**********,7765499**************************
gi 746482918:c54323-54099		ACGUUGAGUGAAGCACCCCUUUA.UGUUGUCaU.AC.AGACCUGUUUUGACGCCUGCCc-CUUaACcgGGCAGGCGUUUUUUU
#=GR gi 746482918:c54323-54099	PP	***************************************
gi 157368249:197999-198236		ACGUUGAGUGAGGCACCGCCC.C.UGUUGUCCU.AGACCUG.AUuqcUUUUUUAUACACUUGCC.ACCGCGGCAAGUGUUUUUUU
#=GR qi 157368249:197999-198236	PP	***************************************
gi 652427325:24161-24383		ACGUUGAGUGAGGCACAAUCC GUC. U. GU. AGAC CUG. A U uG U AUUUUGACACCUAUC. UUU U
#=GR gi 652427325:24161-24383	PP	***************************************
gi 416288147:41498-41717		ACGUUGAGUGAAGCACCCACU.UGUUGUCaU.AC.AGACCUGU.UUUAACGCUUGCUCC.GUAAUAAGAGCAGGCGUUU
#=GR gi 416288147:41498-41717	PP	***************************************
gi 522834473:486-694		ACGUUGAGUGAGCACCGCCC.C.UGUUGUCCU.AGACCUG.AUUg.cuUCUUUAUACACUUGCC.CUGC
#=GR gi 522834473:486-694	PP	***************************************
gi 525689963:162274-162494		ACGUUGAGUGAGCACCGCCC.C.UGUUGUCCU.AGACCUG.AUu.gCUUUUUAUACACUUGCC.ACCGC
	PP	***************************************
gi 739582859:174571-174797		ACGUUGAGUGAGGCACCACCC.C.UGUUGUUCU.AGACCUG.AUUGCUUCUUAUACACUUGCC.UACUC

gi 484142579:c35264-35059		ACGUUGAGUGAGCACCAUCCGUC.U.GU.AGACcugaUUG.U.AUAUUUUAUGCACCUGUA.UUUUu.uAAUACGGGUGUUUUUU-
#=GR gi 484142579:c35264-35059	PP	********986**.*.***9876568.999***********9.876633569**********
gi 739068459:c153565-153344		ACGUUGAGUGAGCCACCAUCCGUC.U.GU.AGACc.ugA.UugUAUAUUUAUACGCCUGUA.UAGUuuUAUACAGGCGUUUUUUU
	PP	********986999.9.9995.432.2113567899*********9.865433779*************
gi 636834654:c47874-47649		ACGUUGAGUGAAGCACCCAAU.U.UGUUGUC.AaAC.AGACCUGU.UUUAACGCCUGCUCU.GGUUUCAGAGCAGGCGUUUUU
#=GR gi 636834654:c47874-47649	PP	***************************************
gi 377579417:c44476-44277		ACGUUGAGUGAAGCACCUAUU.UGUUGUCAC.AGACCUCUUUAACACCUGCC.UUAU
	PP	***************************************
gi 422808133:38920-39144		ACGUUGAGUGAAGCACCCAAU.UGUUGUC.AmaC.AGACCUGU.UUUAACGCCUGCUCC.GUAAUAAGAGCAGGCGUUUUU
	PP	***************************************
gi 746446841:36317-36544		ACGUUGAGUGAGCAC-UAUU.C.UGUUGUC.U.GAcAGACCUG.AUAUUUU-CAACGCUUACC.GUUUaUC
#=GR gi 746446841:36317-36544	PP	***************************************
gi 745755726:c51337-51115		ACGUUGAGUGAAGCACCCAUU. U. AGUUGUC AA. AGAC
#=GR gi 745755726:c51337-51115	РΡ	***************************************
gi 489108174:c48832-48607	-	ACGUUGAGUGAAGCACCCAUU. UaUGUUGUC AU ACA. G A C C UGUUUAACGCCUGCU CCUU aCC
	PP	********9874. 214444444
gi 157144296:136728-136949		ACGUUGAGUGAAGCACCCAUU. UaUGUUGUCaU. AC. AGAC
#=GR gi 157144296:136728-136949	rP	**************9875.32566666644.45.5666
#=GC SS_cons #=GC RF		<pre><<>,))))))),</pre>
#-UC RF		aconomouoaceccarcarcarcarcarcarcarcarcarcarcarcarcar

Postprocessing and Method comparsion

► CMcompare - finding existing models

- ► Rfam as Gold standard
- ► Comparison of search results
- ► Thresholds and false positives

Thanks

▶ Thanks for funding



- ► Thanks to Ivo, Christian
- ► Thanks to my PhD-comitee: Ivo L. Hofacker Peter F. Stadler Renée Schroeder
- ► Thanks to my coworkers!
- ► Thanks for your attention!