

(1) Sequences with start and end annotations found by HMMs (39%)

Verification with a structural motif (5.8S/28S hybridization)

- correct hybridization: $p = 1$
- false secondary structure: $p = 0$

Sequences chosen at random:

AF158726 PAU63164 AF444512 AY826766 DQ987342 AY644648 AJ748107 AF322182 Z81435
X78123 EF646765 DQ200927 AF050240 AF335439 AY559123 DQ310533 AF454078 AF210661
AY144535 AJ271061 AF216579 AJ616317 EF558614 AF416411 AF141165 X96844 AY772813
AF534801 AY178419 AF239358 EF635382 AY789278 DQ217516 AF436781 AF239363 Z81425
AY904059 X74921 AF217801 AY460443 EU367452 EF061910 AF440241 AJ621546 AY014155
Z73770 AY396429 AF013231 AF411022 AJ419914 AM404343 AY254541 ECU57666 AF432179
X86892 AF405547 AY210807 AY014102 AY460462 DQ521290 AF126906 AY260974 DQ866627
DQ787386 EU816709 EF670385 Z78514 SFU28059 Y08746 AF498289 EU753730 AF289763
AJ318958 Z78505 AF087488 AF141613 AY268185 AF202418 EU328719 U31998 AY014069
NVU57673 AF239369 EU669110 EF079386 AF113762 AF293694 AY116085 AJ410756
EU483072 AF508826 AF318758 AF218261 AF508837 AF220979 EF189142 EF026604
AY571042 DQ672548 AY546470

(2) Sequences with only start annotation found by HMMs (45%)

Verification with manual check on GenBank for 28S

- 28S missing: $p = 0.86$
 - 28S present, but found with improved search¹: $p = 0.14$
-
- > 28S identified by HMMs: $p = 1$
 - 28S not identified by HMMs: $p = 0$

Sequences chosen at random:

EF153094 AF465846 AF262915 AF046978 AJ577257 AY014070 AF323273 AY796287
DQ539588 AF348048 AF440256 EU314903 AY625622 DQ444241 AY728272 AF348071
AF019874 AJ492809 LKU56778 AJ584756 AJ867057 AY722781 DQ217523 AY911209
AF007471 AY008579 AF074922 DQ357577 AF224361 AF257918 AJ402160 AF374885
AF430255 AY826304 DQ494678 AF074468 AY389040 AF263403 AF531621 AY561254
AF272252 AF074260 AB329691 AF169847 AY911252 DQ782845 DQ642006 AY425627
DQ317346 AF318620 DQ499105 EF153073 EF514642 AF019862 AF460201 AF373089
AF388622 AY781218 EF153069 DQ667248 AY510157 AY907544 AJ459227 AY187225
AF374913 AM159531 EF517832 EF538416 AF467049 AF537443 AY017396 AY911206
EU910157 AJ421950 DQ463340 DQ537484 DQ811773 AY047069 AF218884 AF508814
AY360256 AY591856 EF595860 EF457725 AY624335 AY826313 AF429275 AY251068
EF093528 EU041815 AY911232 AJ582946 AY826236 AY905431 AF262910 DQ518389
PCU16200 AF019870 AJ000122 AF118521

¹ For the improved search, several manipulations were allowed that are not recommended for an automated annotation and must be manually verified: (a) usage of taxon specific HMMs, (b) ITS2 sequences below 150 nucleotides, (c) reverse complementary and (d) acceptance of higher E-values. All sequences resulting from an improved search were manually verified.

(3) Sequences with only end annotation found by HMMs (3%)

Verification with manual check on GenBank for 5.8S

- 5.8S missing: $p = 0.28$
 - 5.8S present, but found with improved search: $p = 0.72$
-

- > 5.8S identified by HMMs: $p = 1$
- 5.8S not identified by HMMs: $p = 0$

Sequences chosen at random:

AY750707 AY679529 AY999169 AF212171 AF168136 AF303728 AF167350 AJ252136
AF403619 AF343673 AF403617 AF403658 AJ300711 AF342798 EF189209 AY125913
AF303704 AY750705 AB236928 EF680785 AF289779 AF013229 AF315653 AJ311816
AF062606 AY859575 AJ270016 AF144125 DQ903314 AJ011312 AB117928 AF168134
AF399079 AB285506 AY382098 AY382122 AF516154 AF396683 AF513654 AF416415
AF403625 AF403597 AF409121 AF216757 AM234654 EF079400 Y07975 X86897 AY323471
DQ385569 AF168144 AB051231 X86898 X83848 AY210844 EF189921 AY856380 Z78450
AY210845 Y16380 AY999174 AY154739 AB211306 X83831 AY557225 DQ175531 AF168145
AY804242 AY210824 AF409122 AY323466 AJ311820 AF303703 AY210821 AF305555
EF680786 EU124887 AF303702 AF403620 EF362624 AY551143 X83844 AY382099 AY211275
AY114156 AF427515 EF035105 AF293852 EF577164 DQ234270 EF158844 D86515 DQ098898
EF560183 AJ286804 AF416414 AF188045 AM398199 AY737455 DQ385593

(4) Sequences with no annotations found by HMMs (13%)

Verification with manual check on GenBank for 5.8S

- 5.8S & 28S missing: $p = 0.65$
 - 5.8S / 28S present, but found with improved search: $p = 0.34$
-

- > 5.8S / 28S identified by HMMs: $p = 0.99$
- 5.8S / 28S not identified by HMMs: $p = 0.01$

Sequences chosen at random:

LBU56006 AJ972623 AJ426558 AF077978 WAU74429 AF289041 AJ242091 AF319143
AJ237962 Z98582 AY354211 Y11509 AF358580 AJ294691 AB027152 Z71932 AF319029
AJ577260 AY038219 AJ628304 EF579684 AF496952 AJ300270 AM230630 AY083038
AB046577 AY763913 AY761144 EF122861 AJ296443 AF346644 AF145224 AF177836
AF140796 AF286307 AJ419833 AF073549 AF022721 Z75736 Y12588 AY180151 AF228388
AJ296448 AY747120 AF130825 AB033891 Y08007 AY478404 Z48119 AY083061 AJ421828
AY026715 AF495458 AF009100 AF319012 Z93846 AY125140 AF058880 X85461 AJ491690
Z93890 DQ831302 DQ146403 Z78462 AF284962 AY083037 AF239840 AB087238 AF144124
AF495832 X80575 AJ274993 AF510413 EU515306 Z75722 AY349188 AY083052 EU137743
AY026695 AF009106 Z28417 AF183773 Z96886 AJ389166 AF130834 X93976 Y10801
AF171991 AF228397 AY096027 AF080441 AF395642 AF040035 AH009464 X78761 Z48137
AF319151 Z98566 AF183799 AF105255