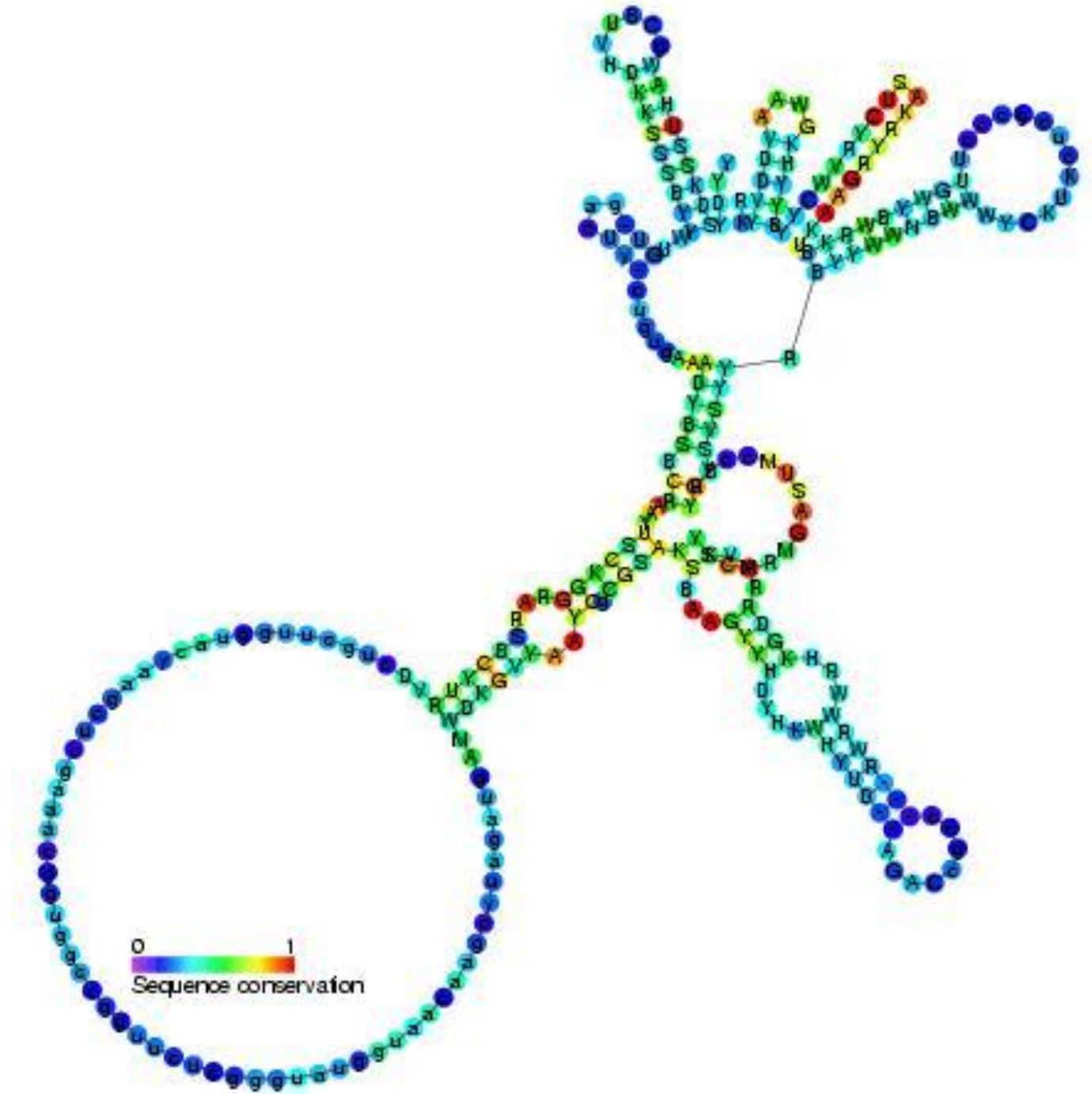


Introns

Group I introns



4801 atctacaaat tctatattggt caatagacct ttgtcaatgg aaatacaatg taattagata
4861 gaaaaaggaa ataaaatagg gacttttgtt ggattggcac gatataaatg cagcaaaaaa
4921 aataggact aagaaagagg caaattgtgt ctaaataatt aaggggtact aaggaccctc
4981 tcctactttt ttattcattt agttcttcaa ttaactcaaa gttctttctt tatctttaaa
5041 gaattctgcc ttccttaaaa tatcataaac agttcttgta ggttgagcac cttttcaag
5101 gaaatagaga atagctggaa catttaaaca agtttgattc tttatcggat cataaaaacc
5161 tacttttcga agatctcttc cttctcttcg agatcgaaca tcaattgcaa cgattcgata
5221 gatagcttat tgggatagat gtagataaac aaccccccc tagaaacgta taggaggttt
5281 tctcctcata cggctcgaga aaatgattcg aatttctgtc tataatgcaa atagatagaa
5341 attcgattat gacttgcatt aatttcctta aagaaaataa aaaacaaatt tcatttatac
5401 tcatgactca agttggctaa ttttgactga caaacttcaa aagaaaaacc cttcgaaatt
5461 ttttgagtcg tctctaaact cttttcttta tctcatctcg aacaaattga cttttattcc
5521 ttattctgat ctaattctat tgttgagacg gttgaaaatc gtgtttactt gttccggaat
5581 cttttatctt tgatttgatg aatccttggg tttagacatt acttcgggaa ctctattct
5641 tttttctttc aaaagagtag caacataccc tttctttttt cttatttcct tcgataaagc
5701 atttcactct tctatagaaa tcaaataatga acaattgatt cagatagact tttaatcaaa
5761 aaagtttttc caatcttcca aaattggact ttttcttatt ttaacctttt gatttcata
5821 ttaaggatag actgacaaag ttggcctaat ttattagttt tcactaacc tagattcttt
5881 cccttgatca aaaagaaatt ctgtccttc gagctccatc gtgtactatt tactttactt
caacc cagcgcaaat tcggttcggg acgaatagaa cagactatgt cgagccaaga
6001 gcattttcat tactatggaa aatgatggat agcaaaatcc acaatcgatc atgtccttca
6061 agtcgcacgt tgctttctac cacatcgttt taaacgaagt tttaccataa cattcctcta
6121 atttcattgc aaaatggat cgagaattga tccaatatgg atggaatcat gaatagtcac
6181 tgcttttgta tactaattca aacctgctat ctatggagaa aattggataa aagaaataag
6241 tatttatcgg qqaacactct qcaaaqatac aatttattta aacctatatt ctttctatca

LibreOffice Impress

```
complement(join(5011  
/gene="rps16"  
/locus_tag="SobiCp000  
/codon_start=1  
/transl_table= 11  
/product="ribosomal p  
/protein_id=" YP_8993  
/db_xref="GI:1186144  
/db_xref="GeneID: 454  
/translation="MVKLRLL  
QTCLNVPAILYFLEKGAQPT
```

BLASTn against that gene as annotated in a closely related genome



(gi|118614470:c6108-6069, c5228-5011) Sorghum bicolor chloroplast, complete genome

Sequence ID: lcl|11559 Length: 258 Number of Matches: 2

Range 1: 37 to 258 [Graphics](#)

[▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Identities	Gaps	Strand
405 bits(219)	2e-115	221/222(99%)	0/222(0%)	Plus/Minus

Query 5012 TTAACTCAAAGTTCTTTCTTTATCTTTAAAGAATTCTGCCTTCCTTAAAATATCATAAAC 5071

Sbjct 258 TTAACTCAAAGTTCTTTCTTTATCTTTAAAGAATTCTGCCTTCCTTAAAATATCATAAAC 199

Query 5072 AGTTCTTGTAGGTTGAGCACCTTTTCAAGGAAATAGAGAATAGCTGGAACATTTAAACA 5131

Sbjct 198 AGTTCTTGTAGGTTGAGCACCTTTTCAAGGAAATAGAGAATAGCTGGAACATTTAAACA 139

Query 5132 AGTTTGATTCTTTATCGGATCATAAAAAACCTACTTTTCGAAGATCTCTTCCTTCTCTTCG 5191

Sbjct 138 AGTTTGATTCTTTATCGGATCATAAAAAACCTACTTTTCGAAGATCTCTTCCTTCTCTTCG 79

Query 5192 AGATCGAACATCAATTGCAACGATTCGATAGATAGCTTATTG 5233

Sbjct 78 AGATCGAACATCAATTGCAACGATTCGATAGATAGCTTATTG 37

Range 2: 1 to 40 [Graphics](#)

[▼ Next Match](#) [▲ Previous Match](#) [▲ First Match](#)

Score	Expect	Identities	Gaps	Strand
75.0 bits(40)	8e-16	40/40(100%)	0/40(0%)	Plus/Minus

Query 6070 GTTGCTTTCTACCACATCGTTTTAAACGAAGTTTTACCAT 6109

Sbjct 40 GTTGCTTTCTACCACATCGTTTTAAACGAAGTTTTACCAT 1

(gi|118614470:c6108-6069, c5228-5011) Sorghum bicolor chloroplast, complete genome

Sequence ID: lcl|11559 Length: 258 Number of Matches: 2

Range 1: 37 to 258 [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
405 bits(219)	2e-115	221/222(99%)	0/222(0%)	Plus/Minus

```
Query 5012 TTAACTCAAAGTTCTTTCTTTATCTTTAAAGAATTCTGCCTTCCTTAAAATATCATAAAC 5071
Sbjct 258 TTAACTCAAAGTTCTTTCTTTATCTTTAAAGAATTCTGCCTTCCTTAAAATATCATAAAC 199
Query 5072 AGTTCTTGTAGGTTGAGCACCTTTTCAAGGAAATAGAGAATAGCTGGAACATTTAAACA 5131
Sbjct 198 AGTTCTTGTAGGTTGAGCACCTTTTCAAGGAAATAGAGAATAGCTGGAACATTTAAACA 139
Query 5132 AGTTTGATTCTTTATCGGATCATAAAAAACCTACTTTTCGAAGATCTCTTCCTTCTCTTCG 5191
Sbjct 138 AGTTTGATTCTTTATCGGATCATAAAAAACCTACTTTTCGAAGATCTCTTCCTTCTCTTCG 79
Query 5192 AGATCGAACATCAATTGCAACGATTCGATAGATAGCTTATTG 5233
Sbjct 78 AGATCGAACATCAATTGCAACGATTCGATAGATAGCTTATTG 37
```

Range 2: 1 to 40 [Graphics](#)

▼ Next Match ▲ Previous Match ▲ First Match

Score	Expect	Identities	Gaps	Strand
75.0 bits(40)	8e-16	40/40(100%)	0/40(0%)	Plus/Minus

```
Query 6070 GTTGCTTTCTACCACATCGTTTTAAACGAAGTTTTACCAT 6109
Sbjct 40 GTTGCTTTCTACCACATCGTTTTAAACGAAGTTTTACCAT 1
```

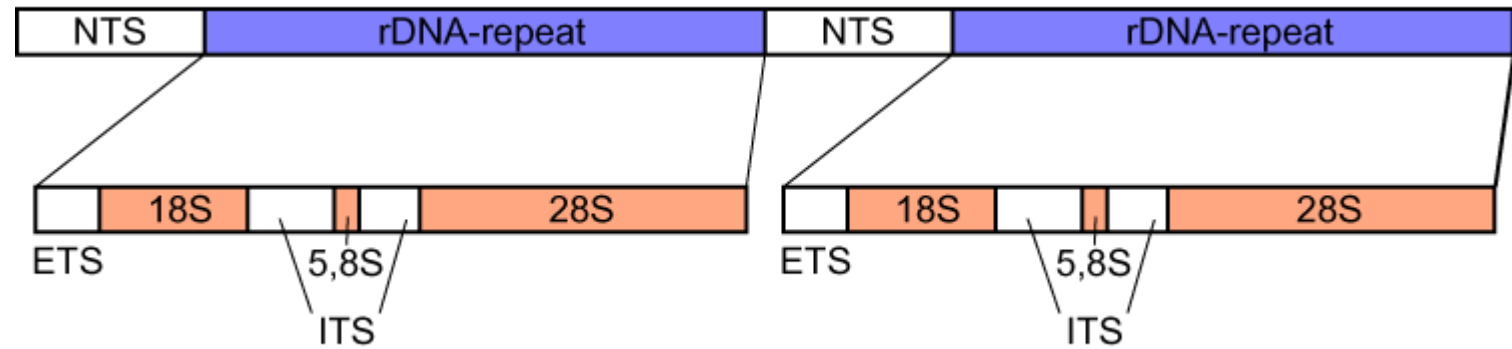
Annotation involves

- What are the positions of the genes
- What do those genes make? (the product)
- Which parts of those genes are exons/introns

- SO, you need to find the boundaries of the key components of the key components of the rDNA

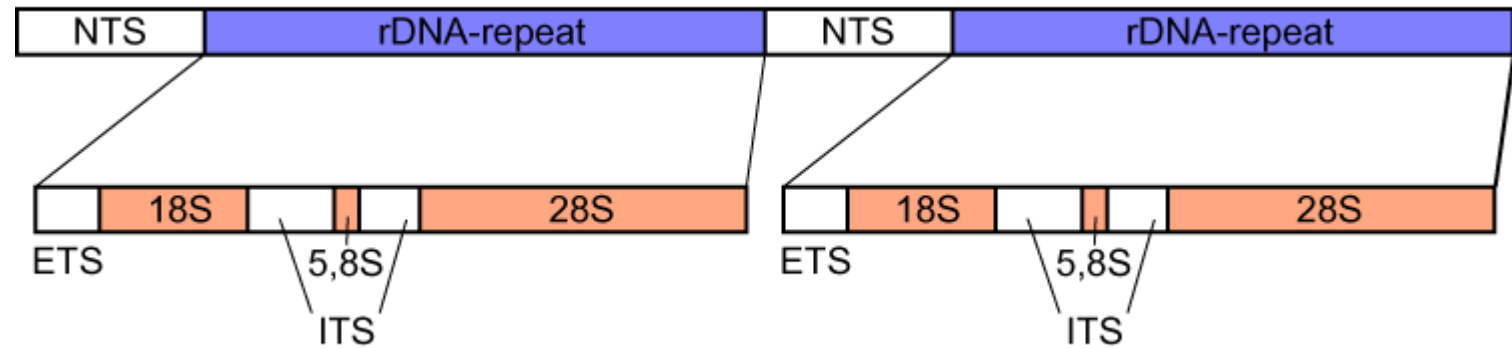
Key components of the rDNA repeat

- ETS
- 18S
- ITS1
- 5.S
- ITS2
- 26S (or 28S)
- ETS



Key components of the rDNA repeat

- ETS
- **18S**
- ITS1
- **5.S**
- ITS2
- **26S (or 28S)**
- ETS



What you need to find

- Beginning of the 18S
- End of the 18S
- Beginning of the 5.8S
- End of the 5.8S
- Beginning of the 26S
- End of the 26S

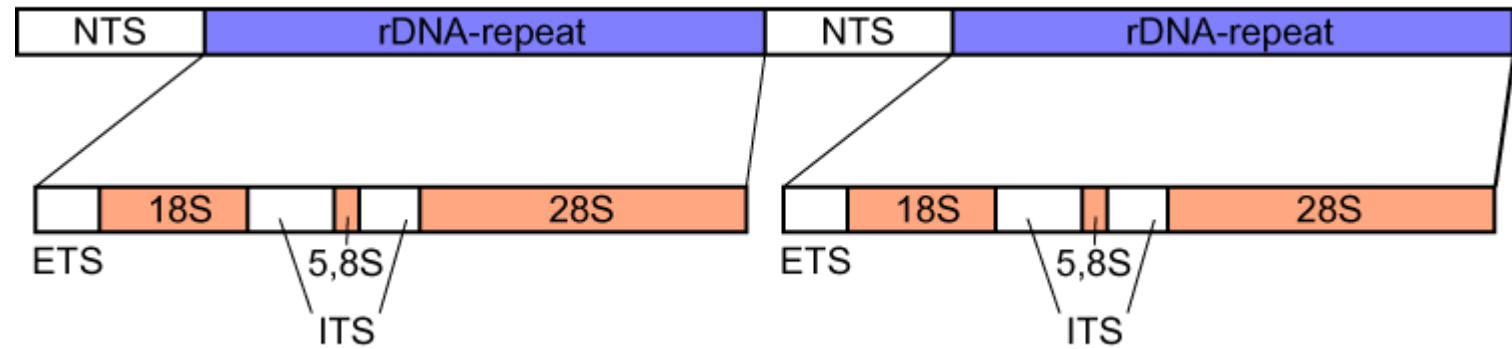
What you need to find

- Beginning of the 18S
- End of the 18S
- Beginning of the 5.8S
- End of the 5.8S
- Beginning of the 26S
- End of the 26S

Incomplete rDNA may not have all of these pieces, in which case you only know that 26S end is $>$ the end of your sequence or 18S starts $<$ the beginning

Key components of the rDNA repeat

- ETS
- **18S**
- ITS1
- **5.S**
- ITS2
- **26S (or 28S)**
- ETS



Also identify whether each piece has introns

- If so, you will have to identify all of the pieces that encode the processed 18S or 26S