## **Find Group of Annotated Regions**

The Find Group of Annotated Regions feature provides an algorithm to search for sequence regions that contain a predefined set of annotations.

Open a DNA sequence in the Sequence View. There are two ways to open the Find Repeats dialog:

- 1. By clicking on the toolbar button:
- 2. By selecting the Analyze Find annotated regions ... context menu item:

	Add			
	Analyze		🕰 Find pattern	ЖF
T	Align	•	SW Find pattern [Smith-Waterman]	ራ羰F
	Cloning	•	🌄 Find ORFs	
	Export		Find annotated regions	
	Edit			

## Algorithm

This tool has been designed to search for annotations that intersect (or completely overlap - it depends on the specified parameters) other, already existing annotations of a given sequence. Let's look at the example:

J_segment (1)								ē	annc	tat	ion	1	$\supset$																													
Loci (1)																													ann	ot	ati	on	2								>	
1 4 6 8 AAAAAAAA	10 12 AAAAA	14 1	5 18 AA	20	22 2 AAA	4 26	28	30	32 AA	34 A A	36 3	38 A A	40 4	2	44 A A	46 A A	48 S	50 S	52 A A	54 A	56 A A	58 A A	60 ( AA	52 ( A A	4 6	6 6	8 7 A A 7	A	AA	74 A A	76 A A	78 A A	80 A A	82 A A	84	86 A A	88 S	90 A A	92 9 AA7	4 9	6 AAA	100 A A

We have a sequence with two annotations. Annotations have different lengths and do not intersect each other. The annotation 2 length is **four** times the an notation 1 length (41 vs. 11 bases).

Using this function, we can find an annotation that intersects both source annotations and captures their shares depending on their lengths. For example, lets find an intersection 25 bases long. We will have the following annotation:

J_s	egmer															ann	iota	ati	on	1	>																														
Loc																					••••••																a	nno	tat	ion	2		_								
mis	c_feat	ure ('																						mi	SC	fe	atu	re																							ï
	AAA	6 AA	8 AA	10 AA	12 AA	14 AA	16 A A	18 A	20 AA	AA	AF	4 2 AA	AA	28 AA	30 A A	32 AA	34 A.A.	36 A A	38 AA	AA	AA	AA AA	46 AA	AA	3 5 AA	AAZ	52 S	54 AA	56 AA	58 AA	60 AA	62 A 7	64 A7	66 A A 7	68 A7	70 A 2	72 AA	74 AA	76 AA	78 A A	80 A A	82 AA	84 AA1	86 AA	88 AA	90 AA2	92 AA	94 AA	96 AAA	10 .AA	A

As we can see, the intersection with the first annotation is **two** characters long, and the intersection with the second annotation is **eight** characters long. This result was chosen because the second annotation is **four** times the length of the first annotation.

J segment (1)	1											
Loci (1)	Ē	1			an	notation	n 2				$ \longrightarrow $	
misc_feature (1)	-	40 [2 bp]	c feature	[8 bp]	63	· · · ·			<b>,</b> , <b>,</b>			
1 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 3 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	8 40 AA	0 42 44 46 48 AAAAAAAAAAA	50 52 54 AAAAAA	56 58 60 62 AAAAAAAA	64 66 68 70 72 AAAAAAAAAAA	74 76 78 AAAAA	AAAA	AAA	36 88 AAAA	90 92 AAAA	94 96 AAAA	AAA
		_										

NOTE: A good candidate for this feature could be any file in <u>Genbank</u> format with a rich set of annotations. <u>FASTA</u> is not the best option, because this format does not store annotations.

## **Parameters**

CDS comment	- [2468, 3467] [3262, 4261]
< <click '+'="" add="" annota<="" button="" new="" th="" to=""><th>ation + [4165, 5164]</th></click>	ation + [4165, 5164]
egion size: 1000 🗘	Save regions as annotations Clear results
O Direct O Complement O E	Both
Annotation must fit into region	

The following parameters are avaliable:

- Left window annotations to search intersection regions for.
- *Right window* the list of possible intersection regions.
- Region size the length of the new intersection region.
- Result strand select the DNA strand whose annotations will be considered in the search. If, for example, the "Complement" strand is selected, but all choosen annotations are on the direct strand, than nothing will be found.
- Annotation must fit into region all annotations, choosen at the left window, must fit completely to the result annotation (completely not just in a few characters). • Save regions as annotations - store results to annotations.
- Clear results clear the result table.