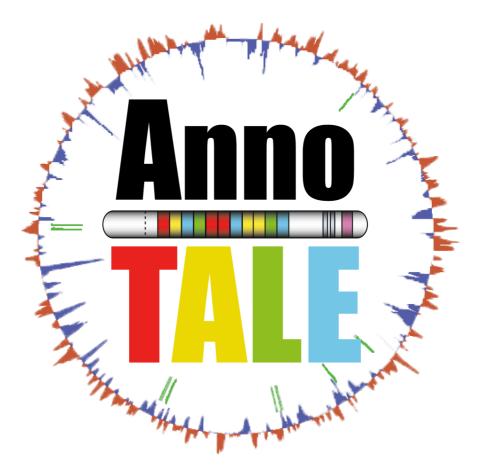
AnnoTALE User Guide

Version 1.0



Predicting, Annotating, and Classifying TALEs from *Xanthomonas spp*.

AnnoTALE User Guide

Maik Reschke, Jens Boch, and Jan Grau

Contact: grau@informatik.uni-halle.de or jens.boch@genetik.uni-hannover.de

AnnoTALE application suite and User Guide can be downloaded at: http://www.jstacs.de/index.php/AnnoTALE

If you use AnnoTALE in your research, please cite:

Jan Grau, Maik Reschke, Annett Erkes, Jana Streubel, Richard D. Morgan, Geoffrey G. Wilson, Ralf Koebnik, and Jens Boch: AnnoTALE: bioinformatics tools for identification, annotation, and nomenclature of TALEs from *Xanthomonas* genomic sequences.

Disclaimer

AnnoTALE is distributed with the intention to be useful, but without any warranty, and without the implied warranty of merchantability or fitness for a particular purpose.

Table of Contents

A Suite of applications - AnnoTALE	3
1) TALE Prediction	5
2) TALE Analysis	6
3) TALE Class Builder (optional)	7
4) Load and View TALE Classes	9
5) TALE Class Assignment	10
6) Rename TALE in File	14
7) Predict and Intersect Targets	15
Quick Start Guide	18
References	19
Download & Installation	19

A Suite of applications - AnnoTALE

AnnoTALE is a program suite for predicting, annotating, and classifying TALEs from *Xanthomonas spp*. The suite consists of seven programs that help the user to analyse a *Xanthomonas spp*. genome sequence with respect to its transcription activator-like effectors (TALEs). AnnoTALE can be used for the prediction of *TALE* genes, classifying TALEs based on their RVD sequences, assigning systematic names to TALEs, and for the prediction of possible target genes in a given target organism (e.g. the rice promoterome).

The interface of AnnoTALE consists of three main parts: the **toolbar**, the **data panel** and the **viewer**. (Fig. 1)

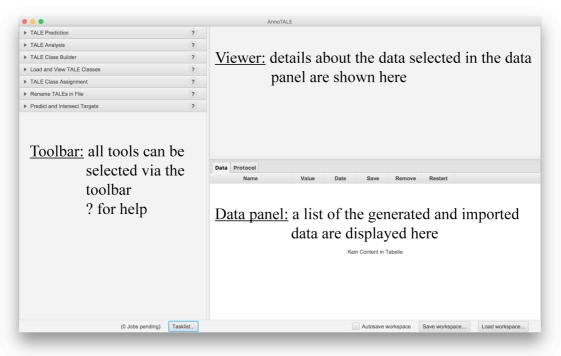


Figure 1: Interface of AnnoTALE after starting the application.

The **toolbar** contains a list of all seven AnnoTALE tools. By clicking on the header of a specific tool, its input parameters are shown and can be specified. The header of each tool includes a button with a ?, which leads to a separate help window with a short description of the purpose and usage of the corresponding tool.

The **data panel** contains two tabs, **Data** and **Protocol**. The **Data** tab shows all data that have been loaded into AnnoTALE or that have been produced as results of one of the AnnoTALE tools. By clicking on one of the entries in the data panel, the corresponding data are visualized in the **viewer** (Fig. 2).

• 🔴 🖷		AnnoTALE										
TALE Prediction	?	>PXO83-tempTALE1 (Pseudo) [2033734- MDPIRSRTPSPAREPLPGPQPDRVQPTADRGVSAP		RRTVSRTRLPSE	PAPLPAFSA	GSSTDRLRPFDPSLF	DTSLFDSMPAVG	PHTEAAP.				
TALE Analysis	?		>PX083-tempTALE2 [2326606-2330017:1] MDPIRSTPSARELJEGPORVOPTAROGROTPRAGGPLGLGLPVRRTMSRTRLPSPAPSPAFSAGSFSDLLRQFDPSLLDTSLLDSMPAVGTPHTAAAI >PX083-tempTALE3 [2331066-2334661:1] MDPIRSTPSARELJEGPORVOPTAROGRAPPAGGPLGLGLPARRTMSRTRLPSPAPSPAFSAGSFSDLLRQFDPSLLDTSLLDSMPAVGTPHTAAAI									
TALE Class Builder	?	>PX083-tempTALE3 [2331006-2334861:1										
Load and View TALE Classes	?	>PXO83-tempTALE4 [2854246-2857561:1										
TALE Class Assignment	?	>PXO83-tempTALE5 [2860837-2864356:1]									
Rename TALEs in File	?	MDPIRSRTPSPARELLPGPQPDRVQPTADRGGAPP >PX083-tempTALE6 [2865345-2868651:1		RRTMSRTRLPSI	PAPSPAFSA	GSFSDLLRQFDPSLI	DTSLLDSMPAVG	PPHTAAAP.				
Predict and Intersect Targets	?	MODTRSRTPSPARELLPGPOPDRVOPTADRGGAPPAGGPLDGLPARRTMSRTRLPSPPAPSPAFSAGSFSDLLROFDPSLLDTSLDSMPAVGTPHTAAA										
		MDPIRSRTPSPARELLPGPQPDRVQPTADRGGAPP	AGGPLDGLPA	RRTMSRTRLPSI	PAPSPAFSA	GSFSDLLRQFDPSLI	DTSLLDSMPAVG	PHTAAAP				
		Name	Value	Date	Save	Remove	Restart					
		PXO83_Genome.gb	FILE		Save	Remove						
		 Result of TALE Prediction (PXO83) 	LIST	Mon Oct	Save all	Remove	Restart					
		GFF: TALE predictions (PXO83)	FILE		Save							
		Genbank: TALE predictions (PXO83)	FILE		Save							
		TALE DNA sequences (PXO83)	FILE		Save							
		TALE protein sequences (PXO83)	FILE		Save							
		TALE pseudo gene matches (PXO83)	FILE		Save							
		 Result of TALE Analysis (PXO83) 	LIST	Mon Oct	Save all	Remove	Restart					
		 Result of TALE Class Assignment (PXO83) 	LIST	Mon Oct	Save all	Remove	Restart					
		 Result of Rename TALEs in File (Genbank) 	LIST	Mon Oct	Save all	Remove	Restart					

Figure 2: Interface of AnnoTALE showing data in the data panel and viewer

Each of the data entries in the data panel may be saved to disk using the Save... button in the corresponding row (Fig. 2). Similarly, sets of results may be saved to a directory using the Save all... button in rows representing aggregate results. In the latter case, the user may specify a target folder and AnnoTALE chooses the names of the output files automatically. In the second tab Protocol, a protocol of current and previous runs of AnnoTALE tools is shown, which contains additional information in case of errors (Fig. 3).

Data Protocol				
	Save protocol			
Starting Load and View TALE Classes				
Downloading current class definition finished.				
Re-building class builder finished.				
Creating class plots for TalAA				
TalAB TalAC TalAD				
TalAE TalAF TalAG				
TalAH TalAI				
TalAJ TalAK TalAL				
TalAL				\sim
	Autosave workspace	Save workspace	Load workspace	

Figure 3: Protocol panel of AnnoTALE

At the **bottom** of the AnnoTALE window (Fig. 2), a bar with further buttons allows the user to access general functionalities. The button Tasklist... shows a list of all running and scheduled jobs of AnnoTALE. Using the buttons Save workspace... and Load workspace...,

the current AnnoTALE workspace, i.e., all data shown in the data panel, may be saved to disk in a monolithic file and loaded from disk, respectively. Loading a workspace from disk adds all data entries stored in that workspace file to the current workspace (instead of replacing the current workspace contents). Using the checkbox Autosave workspace, the workspace will be stored to disk automatically after each and every modification of the workspace, and will also restore that workspace after closing and re-opening AnnoTALE. Currently, this feature should be considered experimental and may reduce the responsiveness of the AnnoTALE GUI.

In the remainder of this User Guide, we will describe each of the seven AnnoTALE tools in more detail.

1) TALE Prediction

The tool TALE Prediction predicts *TALE* genes in a given genomic sequence. With a click on the button Load from file..., a genome sequence (*Xanthomonas spp.*) may be loaded from a file on the local hard drive (Fig. 4). FastA or Genbank files are accepted.

▼ TALE Prediction	?
Genome	
PXO83_Genome.gb 🔻	
Load from file	
The input Xanthomonas genome in FastA or Genbank format	
Strain	
PX083	
The name of the strain, will be used for annotated TALEs (optional)	
Run TALE Prediction	

Figure 4: TALE Prediction – input mask

If a Genbank file is used, all existing annotations will be preserved and annotations for the predicted TALE genes will be added to the existing annotation. It is possible to label the input genome with the name of the *Xanthomonas* strain, in this example PXO83, which is useful when successively predicting TALE genes in several *Xanthomonas* genomes. By clicking on the button Run TALE Prediction..., the prediction tool is started and a progress bar is shown in the lower part of the AnnoTALE window.

• • •		AnnoTALE						
TALE Prediction Genome PX083, Genome ab Classification The input Xanthomonas genome in FastA or Genbank format Strain PX083 The name of the strain, will be used for annotated TALEs (optional)	?	>PR083-temp1ALI: [Penado] 203374-2 >ProspicotrofforeGoccoccocharacteroccoc PR084-temp1ALI2 228466-233007+1 >ProspicotrofforeGoccoccoccocharacteroccoc PR084-temp1ALI2 228466-233007+1 >ProspicotrofforeGoccoccoccoccatecoccoccatecoccoccoccatecoccatecoccoccatecoccoccatecoccoccateoccatecococcatecocatecoccatecoccatecoccateococcatecocatecoccatec	CECGAGCTTCTG CECGAGCTTCTG CECGAGCTTCTG CECGAGCTTCTG CECGAGCTTCTG CECGAGCTTCTG	CCCGGACCCCAI CCCGGACCCCAI CCCGGACCCCAI CCCGGACCCCAI CCCGGACCCCAI	ACCGGATAGI ACCGGATAGI ACCGGATAGI ACCGGATAGI	SGTTCAGCCGAC SGTTCAGCCGAC SATTCAGCCGAC SGTTCAGCCGAC SGTTCAGCCGAC	TGCAGATCGG TGCAGATCGG TGCAGATCGG TGCAGATCGG TGCAGATCGG	GGGGGGGGGCTCC GGGGGGGGGCTCC GGGGGGGGCTCC
Run TALE Prediction TALE Analysis	(?)	ATGGATCCCATTCGTTGGAACAGCCAAGTCCTGCC >PX083-tempTilBE [306002-3810307:1] ATGGATCCCGACGCGGCGGCATCTCCCCCGTGGCCC >PX083-tempTilE9(3811296-3814707:1] ATGGATCCCATTCGTTGGGGAGGCCAAGTCCTGCC >PX083-tempTilE9(381596-3818705:1)	ICGCCIGCGITC CGCGAGCITCIG	TCGGCGGGCAG	CTTCAGCGA	ICTGCTCCGTCA	GTTCGATCCG	TCGCTTCTTGA
 TALE Class Builder 	(?)	Data Protocol						
Load and View TALE Classes	(?)	Name	Value	Date	Save	Remove	Restart	
 TALE Class Assignment 	(?)	PXO83_Genome.gb	FILE		Save	Remove		
Rename TALEs in File	(?)	 Result of TALE Prediction (PXO83) 	LIST	Tue Oct 1	Save a	Remove	Restart	
 Predict and Intersect Targets 	?	GFF: TALE predictions (PXO83) Genbank: TALE predictions (PXO83)	FILE		Save	Kullove	Robart	
		TALE DNA sequences (PXO83)	FILE		Save			
		TALE protein sequences (PXO83)	FILE		Save			
		TALE pseudo gene matches (PXO83)	FILE		Save			
(0 lobs pending) Task	list			Autosave works	pace Sa	ave workspace.	. Load	workspace

Figure 5: Result of TALE Prediction

After the process has finished, Result of TALE Prediction appears in the data panel (Fig. 5). The tool "TALE Prediction" creates a GFF and Genbank file of the genome, where the TALE genes are annotated and labeled with temporary names (tempTALE1...). Furthermore, the program provides FastA files with the DNA sequence, protein sequence and pseudo gene matches of the TALEs in the genome. These data may be used to further analyze the predicted TALEs. GFF and Genbank are standard file formats, which may also be imported into other programs like genome browsers.

2) TALE Analysis

The TALE Analysis tool uses as input a set of complete TALE DNA sequences and splits each of these into the N-terminal region, the individual repeats, and the C-terminal region. Splitting the TALEs into repeats is especially useful to distinguish standard repeats from aberrant, short or long, repeats. It also helps to identify those codons in a repeat that code for

 TALE Analysis 	?
Name	
PXO83	
A name for this run of TALE Analysis (optional)	
PXO83_Genome.gb	
✓ TALE DNA sequences (PXO83)	
TALE protein sequences (PXO83)	
TALE pseudo gene matches (PXO83)	
Run TALE Analysis	

Figure 6: TALE Analysis input mask

its RVD and is, consequently, used to determine the sequence of RVDs for each of the input TALEs. To analyze the TALEs, select the created data set TALE DNA sequences created by the TALE prediction tool, and click the Run TALE Analysis... button (Fig. 6).

The program splits the TALEs of the strain into parts, and creates lists with these parts as a DNA and protein sequence. In addition, the tool automatically extracts the RVD sequence of each TALE (Fig. 7). Further tools in the AnnoTALE pipeline may use the TALE DNA parts or TALE Protein parts output of the TALE Analysis tool.

• • •		An	noTALE						
TALE Prediction TALE Analysis TALE Analysis PX083 A name for this run of TALE Analysis (optional) TALE DNA sequences TALE DNA sequences (PX083) ▼ Lead from file The DNA sequences of the TALEs Run TALE Analysis	? ?	>PR003-tempTALE1 (Ps Ns-HD-NG-NG-NG-NG-NG-NG-NG-NG-NG-NG-NG-NG-NG-	HD-HD-HD-NH-DD 2460-B1-330017:1 2460-B1-330017:1 31006-23346611:3 HD-NN-HD-NS-NG-23346611:3 NS-HD-NN-HD-NS-NG-2857561:1 NS-HD-NN-NN-NN- 60837-28643561:1 HD-NT-HD-NN-B6651:1 HD-NT-HD-NN-B6651:1 HD-NT-HD-NN-HG 5145-2864651:1 HD-NT-HD-NN-HG 5145-286470:1 HD-NT-NN-NN-NN-NN- State-100002-3810:007:1 S-S-16C-HD-NG-NK- 11296-3814707:1 H295-814707:1	-NG-HD-NI-HD- NS-HD-N*-NS- -SS-HD-NI-NI- -NI-NI-NG-HD -NN-NI-HD-NN- HD-HI-ND-HD- +HD-HD-HD-N*- -HD-HD-NI-NG-	NG NN-NI-NN-NI- NS-NG NI-HD-NG-NG NN-NI-NN-HD-		IG−NN−HG−N*		
TALE Class Builder	?	Data Protocol							
Load and View TALE Classes	?	Nar	ne	Value	Date	Save	Remove	Restart	
 TALE Class Assignment 	?	PXO83_Genome.gb		FILE		Save	Remove		
 Rename TALEs in File 	্	Result of TALE Predict	on (PXO83)	LIST	Tue Oct 1	Save a	Remove	Restart	
Predict and Intersect Targets	্	 Result of TALE Analysi 	s (PXO83)	LIST	Tue Oct 1	Save a	Remove	Restart	
		TALE DNA parts (PX	O83)	FILE		Save			
		TALE Protein parts (PXO83)	FILE		Save			
		TALE RVDs (PXO83)	FILE		Save			
(0 jobs pending) T	asklist				Autosave works		e workspace	. Load worksp	1200

Figure 7: Result of TALE Analysis – TALE RVD sequences

3) TALE Class Builder (optional)

For routine use you may skip this step and proceed to the next tool - Load and View Classes. This step is *only* necessary if *no* classification of TALEs exists or to build a new hierarchical classification of TALEs for a specific application.

The tool TALE Class Builder is used to build a hierarchy of a given set of TALEs and group them together in classes. It can be used to create a classification of TALEs, e.g. from different pathovars of *Xanthomonas spp*. The tool groups the TALEs into classes on the basis of an RVD sequence comparison. The tool determines a mismatch score from the pairwise alignment of each pair of two TALEs RVD sequences. Afterwards the TALEs are grouped together such that the average mismatch score in a class does not exceed the user-specified threshold (for detailed information see Grau *et al.* 2015). This class definition can be used as a basis for assigning new TALEs to one of these classes using the tool TALE Class Assignment.

To create your own TALE classes load TALE DNA or protein sequences from a local FastA file or from the output of the TALE Prediction tool or use the previously generated TALE DNA parts or TALE Protein parts from the "TALE Analysis" tool and click on the Run TALE Class Builder... button. If creating a new classification, e.g. for a different pathovar, it is possible to use the default settings for cutoff (5.0) and significance level (0.01) or to customize them depending on the specific application. After finishing the classification the Class builder file, a tree of all classes, as well as a report for the individual classes are shown in the data panel (Fig. 8).

+ - Fit						
Data Protocol						
Name	Value	Date	Save	Remove	Restart	
 Result of TALE Class Builder 	LIST	Tue Oct 1	Save a	Remove	Restart	^
Class builder	FILE		Save			
Tree of classes						
1100 01 0100000	IMAGE		Save			
Class 1	IMAGE LIST		Save			
Class 1	LIST		Save a			
Class 1Class 10	LIST LIST		Save a Save a			
 Class 1 Class 10 Class 11 	LIST LIST LIST		Save a Save a Save a			
 Class 1 Class 10 Class 11 Class 12 	LIST LIST LIST LIST		Save a Save a Save a Save a			

Figure 8: Result of TALE Class Builder - Tree of classes

In the class report of every individual class, information is collected, e.g., the name of the class, RVD sequences of the TALEs in this class, the most likely common binding sequence, the alignment scores, the related classes with significant matches and an RVD sequence alignment of the TALEs from these classes (Fig. 9).

Class 1 for (AffineCosts, 5.0, SEMI_GLO					
distance: -∞	,				
significance: p=0E0					
NI HG NI NI NS HD NN HD HD HD NS N* N'	* HD HD NS	NS NN NN NI	NG NN NI	N* NS N*	PXO83-tempTALE18
Most likely common binding sequence:					
ATAACGCCCACC	C C A	A G G A	T G A	C A C	
Class tree:					
PXO83-tempTALE18)					
lignment scores:					
elated classes with significant matche	es:				
lass 15 with members					
XO83-tempTALE19 related to PXO83-temp?	FALE18 with	score 14.5	(p=6,25E-	-3)	
lignments:					
2X083-tempTALE19 vs. PX083-tempTALE18:					
NI HG NI NI NS HD NN HD HD HD NS N* N'					
:: : : : NI HG NI NI HG HD NN HD HD HD NI NI NI					N74
Cost: 14.5	N NI HD HD	HD HG NN ND	HD NO NN	HD N. NO	N
Data Protocol					
Name	Value	Date	Save	Remove	Restart
Class builder	FILE		Savo		
			Save		
Class builder Tree of classes	FILE IMAGE		Save		
Tree of classes	IMAGE		Save		
Tree of classes Class 1	IMAGE LIST		Save Save a		
Tree of classes Class 1 Class report for 1	IMAGE LIST FILE		Save Save a		
Tree of classes	IMAGE LIST FILE IMAGE		Save Save a Save Save		

Figure 9: Result of TALE Class Builder - Class report

4) Load and View TALE Classes

The tool Load and View TALE Classes loads a given set and hierarchy of TALE classes, termed Class builder into AnnoTALE. Typically, the current definition of TALE classes is downloaded from the server via the Download current definition option. Alternatively, a local file can be loaded into AnnoTALE using the Load from file... button (Fig. 10).

	Load and View TALE Classes	?
(Class definition source	7
	✓ Download current definition	
	Load from local file	requires internet connection) or load
(ienniaon nonniocai nie	ł
-		
	Run Load and View TALE Class	es

Figure 10: Load and View TALE Classes input mask

The local Class builder can be either a class hierarchy that was built using the TALE Class Builder tool before or a saved file from a previous AnnoTALE session (e.g. Augmented Class Builder, output of the TALE Class Assignment tool, see below). After loading, the entry Result of TALE Class Builder appears in the data panel, which consists of the Class builder download file, the tree of classes, and a tree for each single class (Fig. 11).

		AnnoTALE						
TALE Prediction	?	+ - Fit						
TALE Analysis	(?)	TalAB6 Xoo PX086	(tal7b)	NI HG NI NI	NI NN HD	NS NN NS NN H	ID NN NI HD N	
TALE Class Builder	?	- TalAB5 Xoo PX083				NS NN NS NN F		
Load and View TALE Classes	(?)	- TalAB2 Xoo PX095			11 11 11	NS NN NS NN H	T TE TE TE T	
Class definition source		TalAB1 Xoo PX099				NS NN NS NN H	1 11 11 11 1	
Download current definition 👻		TalAB4 Xoo KACCI				NS NN NS NN H		
Download current class definition (requires internet connection) or k definition from local file	ad	2 1.5 1 0.5 0	11019 (2001995) NI HONI NI	NI NN HD	IND NIN ND NIN F		N NI NG HD NG
Run Load and View TALE Classes								
TALE Class Assignment	(?)	<(
Rename TALEs in File	?	Data Protocol						
Predict and Intersect Targets	?	Name	Value	Date	Save	Remove	Restart	
		Result of Load and View TALE Classes	LIST	Thu Oct 1	Save a		Restart	
		Class builder download	FILE			. Remove	Restart	
					Save			
		 Class reports 	LIST		Save a			
		Tree of classes	IMAGE		Save			
		Class TalAA	IMAGE		Save			
		Class TalAB	IMAGE		Save			
		Class TalAC	IMAGE		Save			
		Class TalAD	IMAGE		Save			
		Class TalAE	IMAGE		Save			
		Class TalAF	IMAGE		Save			
(0 jobs pending) Tasl	list			Autosave work		Save workspace		workspace

Figure 11: Result of Load and View TALE Classes - overview

It also contains textual Class reports as well as supplementary files describing the TALEs that are contained in the current class hierarchy and their source strains. The Class reports contain general information about all TALEs that are included in the Class builder, the TALE protein and DNA sequences, and a list of all TALEs and classes (Fig. 12).

#	ID	Accession	Start	End	Strand		
1	TalAA6 Xoo PXO86 (tal5b)	CP007166.1	2541429	2545185	+1		ć
2	TalAA5 Xoo PXO83		2305535	2309291	-1		
3	TalAA2 Xoo PXO99A (tal7b)	CP000967.1	2687333	2691088	+1		
4	TalAA1 Xoo PXO99A (tal8b)	CP000967.1	2899420	2903175	+1		
5	TalAA4 Xoo KACC10331 (XOO2276)	AE013598.1	2401988	2405743	+1		
6	TalAA3 Xoo MAFF311018 (XOO2160)	AP008229.1	2387818	2392653	+1		
7	TalAB6 Xoo PXO86 (tal7b)	CP007166.1	2843503	2847022	+1		
8	TalAB5 Xoo PXO83		2860837	2864356	+1		
9	TalAB2 Xoo PXO99A (tal7a)	CP000967.1	2682825	2686343	+1		
10	TalAB1 Xoo PXO99A (tal8a)	CP000967.1	2894912	2898430	+1		
11	TalAB4 Xoo KACC10331 (XOO2128)	AE013598.1	2226672	2230394	-1		
12	TalAB3 Xoo MAFF311018 (XOO1998)	AP008229.1	2205654	2209376	-1		`
Da	ta Protocol						
	Name	Value	Dat	e Sa	ive Remo	ove Restart	
Ŧ	Class reports	LIST		Save	e a		
	All TALE protein sequences	FILE					
	All TALL protein acquerices	TILL		Save	ə		
	All TALE DNA sequences	FILE		Save	э		
	List of TALEs	LIST		Save	ə		
	List of classes	FILE		Save	ə		
	Tree of classes	IMAGE		Save	ə		
	Class TalAA	IMAGE		Save	ə		
	Class TalAB	IMAGE		Save	ə		
	Class TalAC	IMAGE		Save	ə		

Figure 12: Result of Load and View TALE Classes – List of TALEs

5) TALE Class Assignment

The tool TALE Class Assignment assigns a given set of TALEs, e.g., the TALEs that were previously predicted in a *Xanthomonas* genome, to one of the existing TALE classes that were loaded in the previous tool. If no class with sufficient similarity to an individual TALE exists, this TALE is assigned to its own, new class. The assignment into classes is the basis for the systematic nomenclature of TALEs and the TALE Class Assignment tool will propose systematic TALE names.

▼ TALE Class Assignment	?
Class builder	
Class builder download 🔻	
Load from file	
TALE class builder definition	
TALE sequences	
TALE DNA parts (PXO83)	
Load from file	
The sequences of the TALEs (DNA or protein), or "TALE DNA parts" or "TALE Protein parts" output of "TALE Analysis". Strain	
PX083	
The name of the strain. (optional)	
Accession	
The accesion number of the genome (if applicable). (optional)	
Run TALE Class Assignment	

Figure 13: TALE Class Assignment input mask

As input, the TALE Class Assignment tool needs (i) a given set and hierarchy of TALE classes, the Class builder, and (ii) a set of TALE sequences (Fig. 13). The TALE sequences may be those generated by the TALE Prediction tool, one of the pre-processed TALE parts files produced by the TALE Analysis tool, or DNA sequences of TALEs stored in a local FastA file. In the latter case, these sequences may now be loaded into AnnoTALE using the Load from file... button. Optionally, the user may provide the name of the *Xanthomonas* strain that has been the source of these TALEs and (if available) an accession number, e.g., the accession number of the corresponding genome in NCBI Genbank. If provided, the strain information will be included into the annotation of TALEs with systematic names.

After all parameters have been specified, the assignment is started by clicking on the button Run TALE Class Assignment....

• • •	AnnoTALE					
TALE Prediction ?						
TALE Analysis						
TALE Class Builder						
Load and View TALE Classes						
TALE Class Assignment ?						
Class builder Class builder Class builder TALE class builder definition						
TALE sequences TALE DNA parts (PXO83) V						
Load from file	Data Protocol					
The sequences of the TALEs (DNA or protein), or "TALE DNA parts" or "TALE Protein parts" output of "TALE Analysis".	Name	Value	Date	Save	Remove	Restart
	PXO83_Genome.gb	FILE		Save	Remove	
Strain PXO83	Result of TALE Prediction (PXO83)	LIST	Mon Oct	Save a	Remove	Restart
The name of the strain. (optional)	 Result of TALE Analysis (PXO83) 	LIST	Mon Oct	Save a	Remove	Restart
Accession	 Result of TALE Class Assignment (PXO83) 	LIST	Mon Oct	Save a	Remove	Restart
	▶ Reports	LIST				
The accesion number of the genome (if applicable). (optional)				Save a		
	▶ Classes	LIST		Save a		
Run TALE Class Assignment	TALE names (PXO83)	LIST		Save		
Rename TALEs in File	Augmented class builder (PXO83)	FILE		Save		
Predict and Intersect Targets	 Renamed TALE sequences (PXO83) 	LIST		Save a		
(0 jobs pending) Tasklist			Autosave work	snace S	ave workspace	Load workspace

Figure 14: Result of TALE Class Assignment -overview

signed to class TalAN. her families with significant matche ass TalBZ: distance 10.9, (p=5,56E-4						
Data Protocol			-	-		
Name	Value	Date	Save	Remove	Restart	
Popult of TALE Class Assignment (DVOR2)		Tuo Oot 1				
	LIST	Tue Oct 1	Save a	Remove	Restart	
 Result of TALE Class Assignment (PXO83) Reports 	LIST	Tue Oct 1	Save a Save a	Remove	Restart	
		Tue Oct 1		Remove	Restart	
▼ Reports	LIST	Tue Oct 1	Save a	Remove	Restart	
 Reports Report for PXO83-tempTALE1 (Pseudo) 	LIST FILE	Tue Oct 1	Save a Save	Remove	Restart	
 Reports Report for PXO83-tempTALE1 (Pseudo) Report for PXO83-tempTALE2 	LIST FILE FILE	Tue Oct 1	Save a Save Save	Remove	Restart	
Reports Report for PXO83-tempTALE1 (Pseudo) Report for PXO83-tempTALE2 Report for PXO83-tempTALE3	LIST FILE FILE FILE	Tue Oct 1	Save a Save Save Save	Remove	Restart	
Reports Report for PXO83-tempTALE1 (Pseudo) Report for PXO83-tempTALE2 Report for PXO83-tempTALE3 Report for PXO83-tempTALE4	LIST FILE FILE FILE	Tue Oct 1	Save a Save Save Save Save Save	Remove	Restart	
Report for PXO83-tempTALE1 (Pseudo) Report for PXO83-tempTALE2 Report for PXO83-tempTALE3 Report for PXO83-tempTALE4 Report for PXO83-tempTALE5	LIST FILE FILE FILE FILE	Tue Oct 1	Save a Save Save Save	Remove	Restart	

Figure 15: Result of TALE Class Assignment - Class assignment report

In the data panel, section Result of TALE Class Assignment reports, classes, TALE names, and Augmented class builder are listed (Fig. 14). The reports give information about the input TALEs in the viewer window, e.g. which class the TALE has been assigned to and possible other families with significant matches (i.e. other families with a similar RVD sequence, Fig. 15). In the section Classes all classes are listed that have been modified by the addition of new

TALEs or that have been created, because a TALE did not fir into any of the existing classes. For each of those classes, the tool creates a report and a tree of the members.

	AnnoTALE						
TALE Prediction ?	Class TalAA for (AffineCosts, 5.0, SE distance: 3.85	MI_GLOBAL)					
TALE Analysis	significance: p=8,3E-12						
TALE Class Builder ?	NI HG NS HG HG HD NS NG HD NN NG HG NI HG NS HG HG HD NS NG HD NN NG HG						TalAA2 TalAA1
Load and View TALE Classes	NI HG NS HG HG HD NS NG HD NN NG HG	NG HD HG H	D HD NI NN N	IG			- TalAAS
▼ TALE Class Assignment ?	NI HG NI NG HG HD NS NG HD NN NG HG NI HG NI NG HG HD NS NG HD NN NG HG					HD HD N* NS N	TalAA4 V* TalAA3
Class builder	Most likely common binding sequence:						
Class builder download 🔹	TATATTCATCGTT	тстс	CAGI	CAC	AAA	сссас	:
Load from file							
TALE class builder definition	Class tree: 3.85000000000000 (1.0 (
TALE sequences	0.0 (
TALE DNA parts (PXO83)	< [>
Load from file	Data Protocol						
The sequences of the TALEs (DNA or protein), or "TALE DNA parts" or "TALE Protein parts" output of "TALE Analysis".	Name	Value	Date	Save	Remove	Restart	
Strain	 Result of TALE Class Assignment (PXO83) 	LIST	Mon Oct	Save a	Remove	Restart	
PXO83	Reports	LIST		Save a			
The name of the strain. (optional)	▼ Classes	LIST		Save a			
Accession	Modified class TalAA	LIST		Save a			
	Class report for TalAA	FILE		Save			
The accesion number of the genome (if applicable). (optional)	Class tree for TalAA	IMAGE		Save			
	Modified class TalAB	LIST		Save a			
Run TALE Class Assignment	Modified class TalAC	LIST					
Rename TALEs in File ?				Save a			
Predict and Intersect Targets	Modified class TalAD	LIST		Save a			
	Modified class TalAE	LIST		Save a			
(0 jobs pending) Tasklist			Autosave work	space Sa	ve workspace	Load wo	rkspace

Figure 16: Result of TALE Class Assignment - Class report

The class report (Fig. 16) gives information about the RVD sequences of the TALEs and the alignment score for each TALE pair in the class, the significance of the class assignment and the most likely common binding sequence of all TALEs in this class.

- Fit TaLAA2 Xoo PX099A (tal7b) TaLAA1 Xoo PX099A (tal8b) TaLAA1 Xoo PX093A (tal8b) TaLAA3 Xoo MAFC31031 (X002 TaLAA3 Xoo MAFF311018 (X00	NI HĠ NS HĠ HĠ NI HG NS HG HG HC : : 276) NI HG NI NG HG HC	NS NG HD NN N NS NG HD NN N NS NG HD NN N	IG HG NG HD HG	HD HD NI NN NG HD HD NI NN NG	 HD NI N* NS NI	
Protocol						
Name	Value	Date	Save	Remove	Restart	
 Modified class TalAA 	LIST		Save a			
 Modified class TalAA Class report for TalAA 	LIST FILE		Save a Save			
Class report for TalAA	FILE		Save			
Class report for TalAA Class tree for TalAA	FILE		Save			
Class report for TalAA Class tree for TalAA Modified class TalAB	FILE IMAGE LIST		Save Save Save a			
Class report for TaIAA Class tree for TaIAA Modified class TaIAB Modified class TaIAC	FILE IMAGE LIST LIST		Save Save Save a Save a			
Class report for TaIAA Class tree for TaIAA Modified class TaIAB Modified class TaIAC Modified class TaIAD	FILE IMAGE LIST LIST LIST		Save Save Save a Save a			
Class report for TalAA Class tree for TalAA Modified class TalAB Modified class TalAC Modified class TalAD Modified class TalAE	FILE IMAGE LIST LIST LIST LIST		Save Save a Save a Save a Save a			

Figure 17: Result of TALE Class Assignment - Class tree

The class tree shows the relationship between the individual TALEs of the class and an RVD alignment of the TALEs. The newly assigned TALE is marked in blue (Fig. 17). Furthermore, the program creates a table, TALE names, which shows the temporary and the proposed systematic name of the TALEs.

Name	Value	Date	Save	Remove	Restart	
 Modified class TalAQ 	LIST		Save a			
Modified class TalAR	LIST		Save a			
Modified class TalAS	LIST		Save a			
Modified class TalBA	LIST		Save a			
Modified class TalBD	LIST		Save a			
 Modified class TalBJ 	LIST		Save a			
New class TalCA	LIST		Save a			
TALE names (PXO83)	LIST		Save			
Augmented class builder (PXO83)	FILE		Save			
Renamed TALE sequences (PXO83)	LIST		Save a			

Figure 18: Result of TALE Class Assignment - Augmented class builder

The last file created by the tool is the new class definition termed Augmented class builder (Fig. 18). This file should be saved and used for further TALE class assignments if TALEs from different *Xanthomonas* strains are added to the class hierarchy successively. To ensure that every TALE gets a unique systematic name it is suggested to send the Augmented class builder to Jan Grau (grau@informatik.uni-halle.de) for updating the current definition ("Class builder") on the server. This also ensures that TALE names are reserved, e.g., prior to publication.

6) Rename TALE in File

The tool Rename TALE in File is used to translate the temporary TALE annotation that has been created by the TALE Prediction tool in the *Xanthomonas* genome into the new systematic names produced by the TALE Class Assignment tool.

To rename the TALEs in the GFF or Genbank files from the TALE Prediction tool specify the TALE names file created by the TALE Class Assignment tool and the GFF or Genbank file created by the TALE Prediction tool, e.g., Genbank: TALE Predictions and run the process by clicking on Run Rename TALEs in File. The output is either a GFF or Genbank file, depending of the format of the input file, with the annotated *TALE* genes being renamed from temporary to systematic names. This GFF or Genbank file can be saved and used for further analysis in a genome browser (Fig. 19).

• • •			AnnoTALE						
TALE Prediction	(?)	LOCUS DEFINITION	PX083 start=3214	5029651 bp	DNA linea	ar UNA O	5-OCT-2015		
 TALE Analysis 	?	ACCESSION VERSION	urn:local:.:14440	45418708.4					
 TALE Class Builder 	?	KEYWORDS FEATURES	•	Dualifiers					
Load and View TALE Classes	?	CDS	2033735	2036749					
 TALE Class Assignment 	?	mRNA	2033734						
▼ Rename TALEs in File	?			1AI3 PXOB3" tative pseudo ge	ene"				
Rename Table		CDS	2326607 /gene="Ta	2330017 1CA1 PX083"					
TALE names (PXO83) 🔻		mRNA	2326607						
Load from file		CDS	2331007	2334861					
A tab-separated table containing the old name in the first column and new name in the second column. Output "TALE names" of "TALE Cla Assignment" tool.		mRNA	2331007	1AN3 PXO83" 2334861 1AN3 PXO83"					
Input file		Data Prot	ocol						
Genbank: TALE predictions (PXO83) 💌			Name	Valu	e Date	Save	Remove	Restart	
Load from file		PXO83_G	enome.gb	FILE		Save	Remove		
The input Genbank or GFF3 file that should be renamed.		Result of T	ALE Prediction (PXO83)	LIST	Tue Oct 1	Save a	Remove	Restart	
		Result of TALE Analysis (PX083) Result of TALE Class Builder Result of Load and View TALE Classes			Tue Oct 1	Save a	Remove	Restart	
Run Rename TALEs in File					Tue Oct 1	Save a	Remove	Restart	
Predict and Intersect Targets	?				Tue Oct 1	Save a	Remove	Restart	
		Result of T	ALE Class Assignment (P	XO83) LIST	Tue Oct 1	Save a	Remove	Restart	
		 Result of F 	tename TALEs in File (Ge	nbank) LIST	Tue Oct 1	Save a	Remove	Restart	
		Rename	d Genbank: TALEs	FILE		Save			

Figure 19: Result of Rename TALEs in File

7) Predict and Intersect Targets

The tool Predict and Intersect Targets is used to facilitate an initial analysis of putative target genes of all the individual TALEs from a *Xanthomonas* strain and putative common target genes of a class of TALEs. The tool is based on the statistical model of TALgetter (Grau et al. 2013), but is limited to the top 100 target site predictions and does not calculate p-values for single target sites.

For a detailed analysis of putative target sites it is advisable to use TALgetter available at http://www.jstacs.de/index.php/TALgetter (Command line application) and http://galaxy.informatik.uni-halle.de (Web application).

For predicting possible target sites of TALEs of a *Xanthomonas* strain in a given sequence e.g., the host plant genome or promoterome, it is necessary to load the sequence, that should be analyzed, from a local FastA file via the Load from file... button in the program. For this example the promoterome of rice (*Oryza sativa*) is used.

Predictions can be done for TALEs in FastA or for the Class Builder itself (i.e., all TALEs). Using TALEs in FastA, it is advisable to load Renamed TALE DNA sequences or Renamed TALE protein sequences, because in these data sets the TALEs have the systematic names instead of temporary names. After loading all necessary data click the Run Predict and Intersect Targets... button (Figs. 20 and 21).

AnnoTALE - User Guide

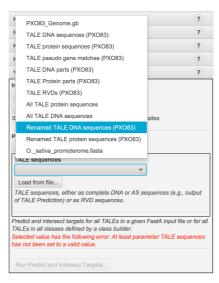


Figure 20: Predict and Intersect Targets input mask, choose renamed TALE sequences

• • •		AnnoTALE						
TALE Prediction	?							
TALE Analysis	?							
 TALE Class Builder 	?							
Load and View TALE Classes	?							
TALE Class Assignment	?							
Rename TALEs in File	?							
 Predict and Intersect Targets 	?							
O_sativa_promoterome.fasta Load from file Sequences, e.g., promoters, to scan for TALE target site	s							
Predictions for		Data Protocol						
TALEs in FastA 👻		Name	Value	Date	Save	Remove	Restart	
TALE sequences Renamed TALE DNA sequences (PXO83) ▼ Load from file TALE sequences, either as complete DNA or AS seque of TALE prediction) or as RVD sequences. Predict and intersect targets for all TALEs in a given Fas TALEs in all classes defined by a class builder. Run Predict and Intersect Targets		 PX083_Genome.gb Result of TALE Prediction (PX083) Result of TALE Analysis (PX083) Result of TALE Class Assignment (PX083) Result of Rename TALEs in File (Genbank) O_sativa_promoterome.fasta 	FILE LIST LIST LIST FILE	Mon Oct Mon Oct Mon Oct	Save a Save a Save a Save a Save a	Remove Remove Remove Remove Remove	Restart Restart Restart	
(0 jobs pendin	g) Tasklist			Autosave work	space Sa	ve workspace	Load v	vorkspace

Figure 21: Predict and Intersect Targets - input mask

The tool generates a data set Overlapping target sites, where putative target sites, which may be predicted to be targeted by more than one TALE, are listed (Fig. 22). The table in the viewer shows the Sequence ID, the intersection size (i.e. the number of different TALEs which are predicted to target this sequence) and all TALEs with rank and position of the target sites indicated as tuple (rank, position).

Furthermore, for every single TALE, predictions of the top 100 putative target sites are generated and displayed in the data panel and viewer, respectively. After selecting the predictions for a TALE a table of the putative targets is shown in the viewer (Fig. 23). The table displays sequence ID and annotation (if available in the input FastA file), the position of the target site in the sequence, the prediction score, the sequence of the target site, and the

• • •		AnnoTALE						
TALE Prediction ?	#		Sequence ID				Intersection size	TalAI3 F
TALE Analysis (?)	1	Os11g34210.2: aspartyl/glutamyl-tRNA ami	idotransferase s	subunit B, putati	ve, expressed	t t	2	(2,342)
TALE Class Builder ?	2	Os11g34210.1: aspartyl/glutamyl-tRNA ami	idotransferase s	subunit B, putati	ve, expressed	d .	2	(3,342)
	3	Os06g46366.1: zinc finger, C3HC4 type, do	omain containin	g protein, expre	ssed		2	
Load and View TALE Classes ?	4	Os09g36680.1: ribonuclease T2 family dom	nain containing	protein, express	ed		2	
► TALE Class Assignment ?	5	Os11g48040.1: mitochondrial carrier protein	n, putative, exp	ressed			2	
▶ Rename TALEs in File ?	6	Os11g34220.1: expressed protein					2	(19,22)
Predict and Intersect Targets (?)	7	Os12g41110.1: OsCML5 - Calmodulin-relat		2				
Input sequences	8	Os12g41715.1: DEAD-box ATP-dependent Os08g14960.1: receptor-like protein kinase					2	
Osativa_promoterome.fasta 💌	10	Os02g22810.1: retrotransposon protein, pu					2	(5,146)
Load from file	11	Os10g16870.1: zinc knuckle family protein,					2	()
Sequences, e.g., promoters, to scan for TALE target sites	<							>
Predictions for		ata Protocol						
TALEs in FastA 👻		Name	Value	Date	Save	Remove	Restart	
TALE sequences	- I	Result of Predict and Intersect Targets	LIST	Tue Oct 1	Save a	Remove	Restart	
Renamed TALE DNA sequences (PXO83) -		Overlapping target sites	LIST					
Load from file					Save			
TALE sequences, either as complete DNA or AS sequences (e.g., output of TALE Prediction) or as RVD sequences.		Predictions for TalAI3 PXO83 (Pseudo)	LIST		Save			
of TALE Fibilition of as RVD sequences.		Predictions for TalCA1 PXO83	LIST		Save			
Predict and intersect targets for all TALEs in a given FastA input file or for a TALEs in all classes defined by a class builder.	//	Predictions for TalAN3 PXO83	LIST		Save			
ALLS III all Gasses defined by a Gass builder.	_	Predictions for TalAF4 PXO83	LIST		Save			
Run Predict and Intersect Targets		Predictions for TalAB5 PXO83	LIST		Save			
		Predictions for TalBJ2 PXO83	LIST		Save			
		Predictions for TalAP3 PXO83	LIST		Save			
		Predictions for TalAQ3 PXO83	LIST		Save			
(0 jobs pending) Tasklist				Autosave works	page Say	/e workspace.	Load works	space

Figure 22: Result of Predict and Intersect Targets – overlapping target sites

match string, which indicates matching RVD-base combinations (M for initial T; I), non perfect matches (:), and mismatches (m for initial T; x).

4936538324
0305463924
0305463924
9493853641
4552997779
8362737726
1187208289
1723319114
8978800786
2121605031
0139507071

Figure 23: Result of Predict and Intersect Targets - target predictions

For a more global view on TALE targets the full class builder can be used in the Predictions for selector. In this case, the prediction is performed for every TALE that is represented in the class builder. Intersections between the predicted sets of targets are determined for all TALEs in common families, which helps to identify putative common targets (and, hence, functions) of TALEs in different *Xanthomonas* strains. Predictions are grouped by the corresponding class.

Quick Start Guide

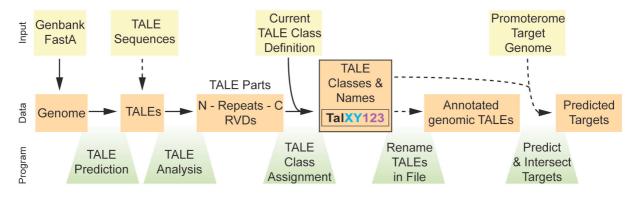


Figure 24: Schematic overview of input parameters, workflow, and program parts.

1) TALE Prediction

Input: Genome file as FastA or Genbank

Output: GFF/Genbank files with annotated TALEs (temporary name); DNA and protein sequences of the TALEs

2) TALE Analysis

Input: TALE DNA or protein sequences from TALE Prediction or as FastA from file **Output:** TALE DNA and protein parts, RVD sequences of all TALEs

3) TALE Class Builder

Input: TALE sequences as FastA, TALE DNA or protein sequences from TALE Analysis

Output: hierarchy and classification of TALEs, Class Builder file, report and tree of the build classes

4) Load and View TALE Classes

Input: Download current definition or Class builder from TALE Class Builder or Augmented Class builder (can also be loaded from local file)

Output: Class builder download, report and tree of the pre-defined classes

5) TALE Class Assignment

Input: Class builder download or Class builder from TALE Class Builder or Augmented class builder and TALE sequences as FastA, TALE DNA or protein parts from TALE Analysis

Output: reports for single TALEs; Modified Classes (report + tree); TALE names; Augmented class builder (for further assignments)

6) Rename TALEs in File

Input: TALE names of TALE Class Assignment or tab separated table and GFF/Genebank file with annotations of TALE Prediction tool

Output: renamed GFF/Genebank file with annotated TALEs with new systematic name

7) Predict and Intersect Targets

Input: sequences to scan, i.e. genome, promoterome as FastA file

Output: predictions of 100 top target sites for all TALEs as table, overlapping target sites of all TALEs

References

Grau, J. *et al.* Computational Predictions Provide Insights into the Biology of TAL Effector Target Sites. *PLoS Comput Biol* 9, (2013).

Grau, J. *et al.* AnnoTALE: bioinformatics tools for identification, annotation, and nomenclature of TALEs from *Xanthomonas* genomic sequences. (2015)

Download and Installation

AnnoTALE is available at http://www.jstacs.de/index.php/AnnoTALE as (i) a runnable Jar file, (ii) a DMG for installation under Mac OS X, and (iii) a Windows installer.

Installation using the runnable Jar file

The runnable Jar file is the preferred version of AnnoTALE if you already have a current version of Java (Java8, update 45 or later) installed on your computer. In this case, no specific installation is required. On each of the three major operating systems (Windows, Linux, Mac OS X), you can just download this Jar file, copy it to any location on your computer you consider appropriate and start AnnoTALE by double-clicking the Jar. If you need to start AnnoTALE with more than the default memory, you need to open a Terminal (or Windows

Command promt), navigate to the directory containing the Jar file, and start AnnoTALE with the larger memory limits, e.g.

java -Xms512M -Xmx6G -jar AnnoTALE-1.0.jar

for allowing AnnoTALE to use 512 MB initially and at most 6GB of RAM.

Installation using the Windows installer

The windows installer comes in three versions with different memory requirements. Depending on the main memory (RAM) that is available on your computer, you should choose the largest version (1GB, 2GB, 6GB) that still fits into your computer's RAM. All standard analysis tasks of AnnoTALE may be executed even with the 1GB version, but large workspaces (e.g., when analyzing several genomes in one AnnoTALE run) may lead to a less responsive GUI or even "out of memory" errors. In the latter case, it may be necessary to restart AnnoTALE.

The windows installer contains an appropriate version of Java in addition to AnnoTALE itself. This version of Java is installed together with AnnoTALE and should not interfere with another Java version already installed on your computer.

After you downloaded the windows installer of choice, start the installation process (Fig. 25). After the installation has finished, you find AnnoTALE in your list of Apps.

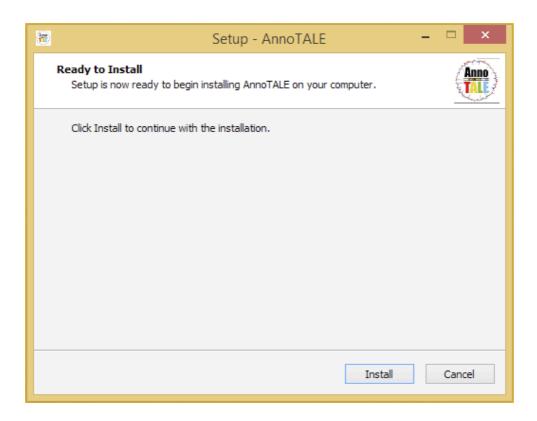


Figure 25: Installation of AnnoTALE using the Windows installer.

Installation from the OS X DMG file

The pre-packaged OS X version of AnnoTALE comes in two version with different memory requirements. Depending on the main memory (RAM) that is available on your computer, you should choose the largest version (2GB, 6GB) that still fits into your computer's RAM. All standard analysis tasks of AnnoTALE may be executed even with the 1GB version, but large workspaces (e.g., when analyzing several genomes in one AnnoTALE run) may lead to a less responsive GUI or even "out of memory" errors. In the latter case, it may be necessary to restart AnnoTALE.

The OS X DMG file contains the AnnoTALE application (Fig. 26), which may be copied to your Applications folder or any other location that you consider appropriate. Afterwards, start AnnoTALE by double-clicking on the AnnoTALE application.



Figure 26: Installation of AnnoTALE for Mac OS X.