User guide – Interpretation of results

(Please Note: Although the following results are taken from the "Browse" section of EffectiveDB, users who upload their own genomic information will receive equivalent results!)

Within the "Browse" section, you see the search box listing genomes and their predicted protein secretion systems. In this case we searched for "Escherichia" (arrow #1 in Fig. 1).

BROWSE EFFECTIVEDB

EffectiveDB provides precalculated results for all 1677 bacterial genomes in the **EggNOG 4.0** database and additional 23 bacterial genomes from the NCBI RefSeq datase. In the near future we will switch to the upcoming EggNOG 5.0 database, derived from STRING 11 (about 10.000 genomes). RefSeq genomes not contained in STRING will be continuously added. We have used the following settings:

- Genome mode: enabled, genotypes from EggNOG, EffectiveS346 enabled (Type III and VI secretion systems only predicted for gram negative bacteria)
- EffectiveT3: enabled if EffectiveS346 predicted a type III secretion system; model 2.0.1
- EffectiveCCBD: enabled if EffectiveS346 predicted a type III secretion system
- T4SEpre: enabled if EffectiveS346 predicted a type IV secretion system, minimal score=0.5
- EffectiveELD: enabled, minimal Z-score=4
- Predotar: enabled (both animal and plant models are used)

Show 10 ⌄ entries   Search: Escherichia ✕ ← 1

Genome	T3SS	T4SS	T6SS
Escherichia albertii TW07627	YES	YES	YES
Escherichia coli 536	NO	YES	YES
Escherichia coli ATCC 8739	YES	YES	YES
Escherichia coli BL21(DE3)	YES	YES	YES
Escherichia coli CFT073	NO	YES	YES
Escherichia coli O157:H7 str. EDL933	YES	YES	YES

Figure 1 Selection of the organism within the "Browse" section. This step obviously is omitted when uploading own data.

The overview immediately shows if there are secretion systems of the three different types available. Clicking on the organims link, leads to the detailed results.

To explain how to interpret the calculated results, we have an in-depth look at the pre-calculated results of Escherichia coli O157:H7 str. EDL933:

To understand which of the calculations are done, we have a closer look at the input parameter, which are depicted at the very bottom of the results page.

PARAMETERS:

```
=========================================================
EFFECTIVE_T3_OLD=selective
EFFECTIVE_T3_NEW=selective
PREDOTAR_ANIMAL_OR_PLANT=a
PROTEINS=INPUT PROTEINS          ←——— 3
MODUS=genome
RUN_PREDOTAR=T
RUN_EFFECTIVE_T3=T    ←——— 2
RUN_T4SEpre=0.5
RUN_EFFECTIVE_CCBD=T
RUN_EFFECTIVE_ELD=4
CheckM_COMPLETENESS=
RUN_EffectiveS346=T
GENOTYPE=COG Profile
EMAIL=
=========================================================
```

Figure 2 Input parameters for the pre-calculated example of Escherichia coli O157:H7 str. EDL933

All of the applications of EffectiveDB, including EffectiveS346 from the genome mode are enabled in this particular case, which can be seen by the capital letter T (=TRUE) (arrow #2 in Fig. 2). In our pre-calculations we used the COG/NOG-profile that was extracted from the EggNOG database. All of the application specific parameters are set to their default values. In detail this means that the results of Effective T3 are selectively chosen by a minimal score of 0.9999. In other words, only hits with very high secretion probability are selected. The results of Predotar are shown in regard to the interaction with animals (arrow #3 in Fig. 2) as hosts. Results for plant-hosts are shown in the output within brackets "()". The check for completeness done by checkM was disabled, as the genomes of the organisms for our pre-calculations are complete (we already checked that on beforehand).

Using the above-described parameters for the calculations leads to the following results (Fig. 3).

EFFECTIVE RESULTS FOR ESCHERICHIA COLI O157:H7 STR. EDL933 (EGGNOG40)

SUMMARY

Number of putative secreted proteins: 956
- Predicted by **EffectiveT3**: 464 (high confidence), 230 (low confidence)
- Predicted by **EffectiveCCBD**: 33 within, 23 before, 32 behind expected region 25..70aa
- Predicted by **T4SEpre**: 101
- Predicted by **EffectiveELD**: 41 eukaryotic-like domains, contained in 125 proteins (40 high, 85 medium Z-score)
Predicted protein secretion systems: Type III, Type IV, Type VI

Figure 1 summary of the calculated results for Escherichia coli O157:H7 str. EDL933

At first, the user gets confronted with an overview of the different calculations. The first line gives us the overall amount of putative secreted proteins. This is not necessarily the same number as summing up the numbers of all of the effector predicting methods. There might be some proteins that can be detected by more than one method. By providing different approaches of effector predictions, we want to find as many putative effectors as possible. Note: not every secreted protein is detected by all methods!

PROTEIN BASED RESULTS

Putative secreted proteins (weak evidence indicated by parentheses)

Show 10 ⌄ entries

Protein	EffectiveT3	EffectiveCCBD	T4SEpre	EffectiveELD	Predotar
155864.Z3464	+	+			
155864.Z5788	+	+			
155864.Z1336	+	(<)			(ER)
155864.Z5450	+	(<)			(plastid)
155864.Z2428	+	(<)			
155864.Z0678	+	(<)			
155864.Z1412	+	(>)			
155864.Z2816	+		+		(mitochondrial)
155864.Z5140	+		+		(mitochondrial)
155864.Z5801	+		+		

Showing 1 to 10 of 956 entries

◄ Previous   Next ►

Figure 2 The example of Escherichia coli O157:H7 str. EDL933 shows that the protein 155864.Z3464 for example, is predicted to be secreted by EffectiveT3 and EffectiveCCBD. The results are ranked by their significance. Proteins that are predicted to be secreted by more than one method are ranked at the top.

Figure 4 shows the protein-based results of the EffectiveDB predictions. Again it can be seen that for instance Protein 155864.Z3464 is predicted not only by EffectiveT3, but also by EffectiveCCBD. Proteins that are not predicted to be secreted are not contained in the table.

To further analyze the diverse results of Effective, the user may also download the combined results as .csv or .html format, listed below the table. Moreover, the specific results of each protein-based prediction can be obtained separately as downloadable files.

Regarding the results of the genome based approach of EffectiveS346, it is first summarized if there are one or several intact secretion systems predicted within the genomic sequence, or not (Fig. 5).

GENOME BASED RESULTS

Summary

- Gram negative: YES (FDR = 0.5 %)
- Functional Type III secretion system: YES (FDR = 12.5 %)
  Estimated copy number = 2.2 (22/10 KEGG proteins)
- Functional Type IV secretion system: YES (FDR = 7.2 %)
  Estimated copy number = 0.0 (0/10 KEGG proteins)
- Functional Type VI secretion system: YES (FDR = 7.7 %)
  Estimated copy number = 1.25 (10/8 KEGG proteins)

Individual results for each secretion system

- List of 100 most important COGs/NOGs for prediction of functional Type III secretion system (T3_top100_COGS.xls)
- List of 100 most important COGs/NOGs for prediction of functional Type IV secretion system (T4_top100_COGS.xls)
- List of 100 most important COGs/NOGs for prediction of functional Type VI secretion system (T6_top100_COGS.xls)
- List of COGs assigned to Type III secretion system in KEGG (T3SS_KEGG_cmp.xls)
- List of COGs assigned to Type IV secretion system in KEGG (T4SS_KEGG_cmp.xls)
- List of COGs assigned to Type VI secretion system in KEGG (T6SS_KEGG_cmp.xls)

Figure 3 Output of genome based calculations of Escherichia coli O157:H7 str. EDL933

In this genome, intact Type III, IV and VI secretion systems have been predicted. Otherwise the prediction would be NO (not available) or N.D. (not defined). The latter might occur within the group of Gram-positive bacteria. Those organisms are known to lack the Type III or VI secretion systems at all and we therefore disable the respective models for Gram-positive bacteria.

Within the brackets, beside the actual prediction, the FDR (=false discovery rate) gives the user an additional indication of the reliability of the result.

The estimated copy number is an additional result, which assigns the actual number of structural proteins (as represented in KEGG) to the significant COGs. This can be an indication of how many secretion systems of this type might have been predicted.

It is very important to conscientiously further investigate these numbers, as they do not directly belong to our actual prediction. Therefore we recommend to use the database T346Hunter. For genomes that are covered by T346Hunter and EffectiveDB, we provide a direct link to their results at the end of our results page.

To do further investigations regarding the provided predictions, users may download all the results in commonly used .csv or .xls format.