FIVA

Blom, E.J.; Bosman, D.W.; van Hijum, S.A F T; Breitling, R.; Tijsma, L.; Silvis, R.; Roerdink, J.B.T.M.; Kuipers, O.P.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in University of Groningen/UMCG research database](#)

Utilization of multiple functional modules in FIVA

An analysis on the overlap between functional annotations was performed (see interpretation in http://bioinformatics.biol.rug.nl/standalone/fiva/data/informationAnalysis.pdf). This analysis shows that the Gene Ontology module is the most informative annotation type for our test organism *Bacillus subtilis*. However, we also demonstrate that adding other annotations provide unique information on top of the GO information. The first information analysis was focused towards the available information content from annotation modules and their overlap. This analysis investigates the results from all annotation modules of a FIVA analysis on two transcriptome datasets (CodY and Mannitol/sorbitol dataset from http://bioinformatics.biol.rug.nl/standalone/fiva/suppMaterials.php).

Several other programs overlap with FIVA in terms of functionality and data sources employed (Khatri, P. et al. (2005). The Gene Ontology annotation module is used most often in these software packages, so this analysis will attempt to demonstrate that combining multiple functional modules into one tool is clearly beneficial to using one or a few modules.

Analysis on datasets

Two functional categories (urea cycle & competence) were found to be significantly overrepresented by functional annotations other than GO (metabolic pathways & interactions) for the CodY dataset. For the carbon sources dataset, five categories were found by the metabolic pathways module which were missed by GO. Below in Table 1A and Table 1B the missed categories by GO are described, together with the genes which were found to be involved in these categories. The genes from the functional categories were checked for relevant GO annotations. Table 1C describes the missed categories by GO and occurrences of relevant functional GO annotations. This table is not based on the genes of our model organism. This table was constructed to identify relevant GO occurrences which are present in the global GO database and describe the missing categories but are missing in the GO annotation for our organism. Furthermore, we checked for perfect GO hits (i.e. GO terms which match the missed category name perfect) but also for GO hits that are involved in the process described by the missed relevant category. We find that in most cases missed categories are available in the GO annotation but not sufficiently annotated. However, in some cases there is no appropriate GO annotation present.

Table 1A. Analysis of functional categories found by functional annotations other than GO (CodY dataset).
This table describes some functional categories that were found overrepresented by the metabolic pathways module (column 1) and the gene involved in these categories (column 2). Relevant GO annotations are displayed in column 3 with their description in column 4. The number of genes that are linked to these GO annotations are indicated in column 5. For example; the metabolic pathway module analysis found an overrepresented category called Urea Cycle in which six genes were linked. Two GO annotations are available that could indicate the similar process. Only three genes are available in this GO annotation, of which two differentially expressed (ureA and ureB).

Original functional category missed by GO	Genes involved in functional category missed by GO	Available relevant GO annotation	Relevant GO description	Total nr of genes with relevant GO
Urea Cycle (from the metabolic pathway annotation)	**argC, argJ, proH, rocD, urea, ureB**	**ureA:** GO:0009039	urease activity	3 genes
		ureB: GO:0009039		
Competence (from the regulatory interactions annotation)	**comC,comFA,comGA comK,nucA,recA, ydbK,yneA,yvyF,ywpH comEA,smf**	**comEA:** GO:0030420	establishment of competence for transformation	2 genes
		smf: GO:0009294	DNA-mediated transformation	1 gene

Although some of the genes contain a relevant GO annotation (Table 1A), it is highly unlikely that they would have been identified in a significant overrepresented GO category due to the small size of the functional GO category. Without the metabolic pathway and regulatory interaction information, the two functional categories described in Table 1 would have been missed.

Table 1B. Analysis of functional categories found by functional annotations other than GO (Mannitol/sorbitol dataset).
For explanation on this table, see Table 1A.

Original functional category missed by GO (all from the metabolic pathways annotation)	Genes involved in functional category missed by GO	Available relevant GO annotation	Relevant GO description present in our model organism.	Total nr of genes with relevant GO
Valine, leucine and isoleucine degradation	**dhaS , mmgA , scoA , scoB , yngE , yngF , yngG , ysiB , ysnE , yusK , yusL , ywnH**			
Fructose and mannose metabolism	**levD , levE , levF , levG , mtlA , mtlD**	**mtlD:** GO:0008926	mannitol-1-phosphate 5-dehydrogenase activity	1 gene
Fatty Acid biosynthesis	**adhB , dhaS , mmgA , yngF , yngI , ysiB , yusK , yusL**	**yusL:** GO:0006631	fatty acid metabolism	3 genes
Butanoate metabolism	**acoB , dhaS , mmgA , mmgB , scoA , scoB , yjmD , yngF , yngG , ysiB , ysnE , yusL , ywnH**			
Benzoate degradation	**mmgA , mmgB , yjmD , yngF , ysiB , ysnE , yusL , ywnH**			

From Table 1B it is clear that quite a number of functional categories that were found using the metabolic pathway annotation module were not detected by the GO annotation. The size of the relevant functional GO categories that were found for the carbon sources dataset would have prevented them to be identified as significantly overrepresented.

Table 1C. Analysis of functional categories found by functional annotations other than GO compared to available GO terms from the global GO database.
This table summarizes the functional categories that were missed by GO. Perfect and second best hits are displayed in column two and three.

Original functional category missed by GO	Available perfect hit from the GO annotation.	Available hit (not perfect) from the GO annotation
Urea Cycle	**urea cycle** GO:0000050	
Competence	n.a.	**competence** pheromone activity GO:0030413 establishment of **competence** for transformation GO:0030420 regulation of establishment of **competence** for transformation GO:0045304
Valine, leucine and isoleucine degradation	n.a.	**isoleucine** metabolism GO:0006549 **isoleucine** catabolism GO:0006550 **valine** metabolism GO:0006573 **valine** catabolism GO:0006574 **leucine** catabolism GO:0006552 **leucine** metabolism GO:0006551
Fructose and mannose metabolism	n.a.	**fructose** metabolism GO:0006000 **fructose** catabolism GO:0006001 **mannose** metabolism GO:0006013 **mannose** catabolism GO:0019309
Fatty Acid biosynthesis	**fatty acid** biosynthesis GO:0006633	**fatty acid** metabolism GO:0006631 **fatty acid** catabolism GO:0009062
Butanoate metabolism	n.a.	n.a.
Benzoate degradation	n.a.	**benzoate** metabolism GO:0018874

From the above table it is evident that a relevant GO category is available for two of the seven missed categories. For the remaining five, one category was not described at all by the GO database (Benzoate degradation) and four were only described by other, more detailed, GO categories.

Concluding, in most cases of categories missed by the GO module the specific *Bacillus subtilis* annotations were missing. In particular, one category identified by the metabolic pathways module would have been missed completely by GO. Furthermore, the question remains whether the four categories, not described by a perfect hit, would have been found at all by GO. The found GO categories could easily be too detailed (no general GO category is available for these categories), and therefore missing the general picture. A solution could consist of adding more general nodes to the Gene Ontology graph. Until then, we conclude that combining multiple functional modules is clearly beneficial to using one module.