

Article Title

First Author¹, Co-Author², Co-Author^{1,3*}

¹ Department X, Institution X, City X, State XX, Country X

² Department X, Institution X, City X, State XX, Country X

³ Department X, Institution X, City X, State XX, Country X

Correspondence*:

Co-Author

email@uni.edu

2 ABSTRACT

Abstract length and content varies depending on article type. Refer to <http://www.frontiersin.org/about/AuthorGuidelines> for abstract requirement and length according to article type.

Keywords: Text Text Text Text Text Text Text Text

1 INTRODUCTION

Hello, how are we doing?

2 METHODS

2.1 Clean up

All functional annotation sets were cleaned up the following way (using definitions from the Gene Ontology version 2019-07-01):

1. Any annotations where the GO accession was marked as obsolete were removed.
2. Some terms in the GO have 'alternative ids'. When naively removing duplicates, two entries will not be recognized as duplicates if they have different accessions pointing to the same GO term. Therefore, all GO accessions were changed to their respective 'main id' and the dataset was again scanned for duplicates.

Table 1 provides information on the number of annotations that were removed this way from each dataset. All further analyses were performed on the cleaned datasets since we assume the user will only be interested in still valid and non-redundant functional annotations.

3 RESULTS

... a quantitative comparison of the datasets in Table.

Table 1. Number of removed annotations during cleanup.

Genome	Dataset	Obsolete Annotations	Duplicates
<i>Glycine max</i>	GOMAP	203	0
<i>Hordeum vulgare</i>	GOMAP	101	0
<i>Medicago truncatula</i> A17	GOMAP	0	0
<i>Medicago truncatula</i> R108	GOMAP	0	0
<i>Oryza sativa</i>	GOMAP	111	2
	GoldStandard	38	556
<i>Phaseolus vulgaris</i>	GOMAP	0	0
<i>Triticum aestivum</i>	GOMAP	285	0
<i>Vigna unguiculata</i>	GOMAP	0	0
<i>Zea mays</i> B73.v4	GOMAP	752	83
<i>Zea mays</i> Mo17	GOMAP	726	77
<i>Zea mays</i> PH207	GOMAP	798	76
<i>Zea mays</i> W22	GOMAP	754	82
<i>Arachis hypogaea</i>	GOMAP	0	0
<i>Zea mays</i> B73.v3	GOMAP	1107	70
	GoldStandard	1	0
	Gramene49	94	2
	Phytozome	54	0
<i>Oryza sativa</i>	Gramene61-IEA	10	14
	Gramene61-all	48	9565

Table 2. Quantitative metrics of the cleaned functional annotation sets. CC, BF, MP, and A refer to the aspects of the GO: Cellular Component, Biological Function, Molecular Process, and Any/All.

Genome	Genes	Dataset	Annotations ^a				Annotated Genes [%] ^b				Median Ann. per G. ^c			
			CC	BF	MP	A	CC	BF	MP	A	CC	BF	MP	A
<i>Arachis hypogaea</i>		GOMAP	153433	132944	493799	780176	57667	56855	67123	67124	2	2	6	10
<i>Glycine max</i>		GOMAP	129215	113827	417555	660597	46020	47034	52871	52872	2	2	6	11
<i>Hordeum vulgare</i>		GOMAP	88130	80282	272823	441235	35237	36470	39733	39734	2	2	5	10
<i>Medicago truncatula</i> A17		GOMAP	107362	99719	364065	571146	42325	43736	50443	50444	2	2	6	10
<i>Medicago truncatula</i> R108		GOMAP	112343	108031	382322	602696	40332	50220	55706	55706	1	2	5	9
<i>Oryza sativa</i>		GOMAP	72780	64685	248700	386165	28619	29853	35824	35825	2	2	6	9
		GoldStandard	7730	11060	19378	38176	5725	7383	9031	11387	1	1	1	3
		Gramene61-IEA	14633	32787	39105	86529	10771	15537	16705	21446	1	1	1	3
		Gramene61-all	20622	40674	54402	115710	13272	16962	18513	22272	1	1	2	4
<i>Phaseolus vulgaris</i>	100	GOMAP	72005	64583	229630	366218	25934	25539	27432	27433	2	2	6	11
<i>Triticum aestivum</i>		GOMAP	267741	218623	785960	1272324	95604	98187	107890	107891	2	2	6	10
<i>Vigna unguiculata</i>		GOMAP	75867	68313	243278	387458	27173	27124	29772	29773	2	2	6	11
		GOMAP	135211	87420	291251	513882	34866	38073	39468	39469	3	2	6	11
<i>Zea mays</i> B73.v3		GoldStandard	1565	65	299	1929	1548	60	151	1634	1	0	0	1
		Gramene49	20072	31056	30089	81217	11834	17991	15800	21926	1	1	1	3
		Phytozome	4787	19044	13100	36931	4524	13728	11365	16132	0	1	1	2
<i>Zea mays</i> B73.v4		GOMAP	88827	82251	278719	449797	36717	37337	39323	39324	2	2	6	10
<i>Zea mays</i> Mo17		GOMAP	87567	79214	277787	444568	33618	35105	38619	38620	2	2	6	10
<i>Zea mays</i> PH207		GOMAP	90617	85500	288677	464794	35170	36762	40556	40557	2	2	6	10
<i>Zea mays</i> W22		GOMAP	95390	85039	289780	470209	36987	37685	40689	40690	2	2	6	10

^a How many annotations in the CC, BF, and MP aspect does this dataset contain? A = How many in total? A = CC + BF + MP^b How many genes in the genome have at least one GO term from the CC, BF, MP aspect annotated to them? A = How many at least one from any aspect? (A = CC ∪ BF ∪ MP)^c Take a typical gene that is present in the annotation set. How many annotations does it have in each aspect? A = How many in total? Ask your favorite statistician why A ≠ CC + BF + MP

Table 3. Quality evaluation of the used GO annotation sets.

Genome	Dataset	SimGIC2 score
<i>Oryza sativa</i>	GOMAP	0.248394
	GoldStandard	0.999974
	Gramene61-IEA	0.319345
	Gramene61-all	0.732175
<i>Zea mays</i> B73.v3	GOMAP	0.024565
	GoldStandard	0.996264
	Gramene49	0.043521
	Phytozome	0.021982