SMTB

ZIMIN FOUNDATION

Petrovax

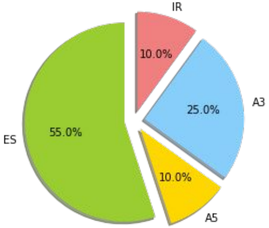
ИНСТИТУТ БИОИНФОРМАТИКИ

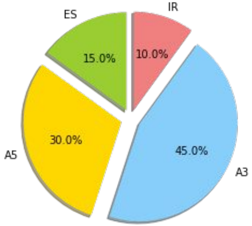
Alternative splicing universe

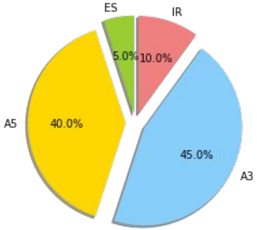
Students: Andrey Kostarev, Yaroslava Mikhailenko
Supervisors: Olga Tsoy, Grisha Ryabykh, Liubov Lonishin

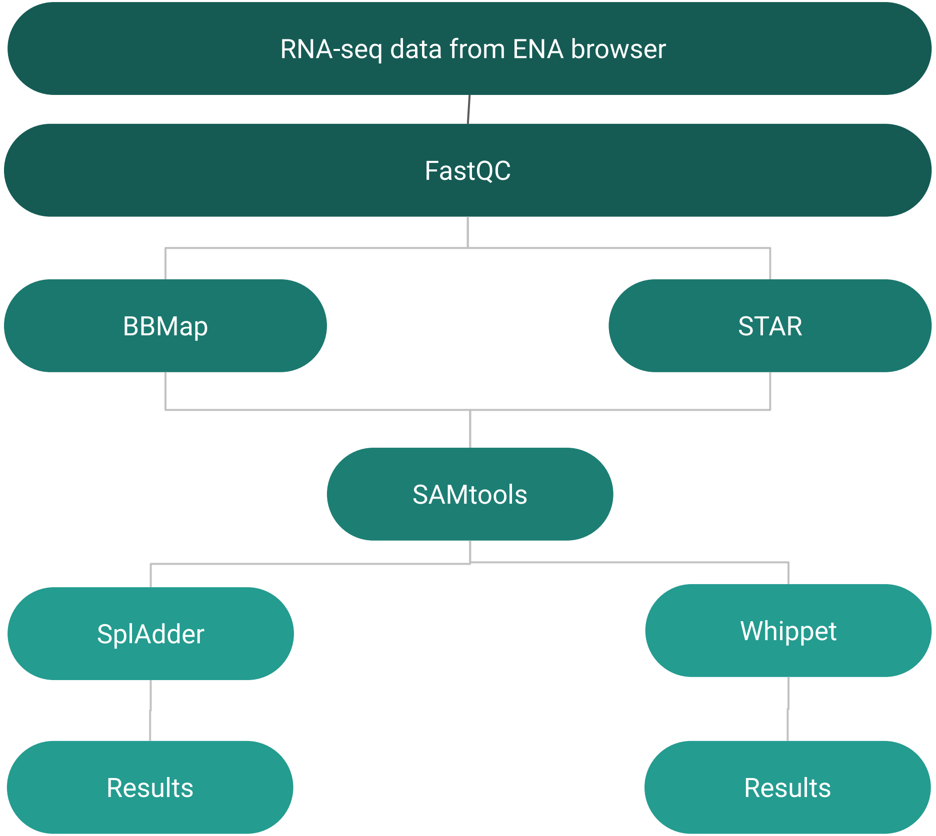
Motivation


AVENGERS, ASSEMBLE!
choose your favorite superhero
Strength
Skill
Brains
Gadgets
Heart
Captain America
Greatest Strength Tactics
Greatest Weakness Human
Swagger Meter 50%
The Hulk
Greatest Strength Rage
Greatest Weakness Rage
Swagger Meter 5%
Iron Man
Greatest Strength Suit
Greatest Weakness Battery
Swagger Meter 110%
Thor
Greatest Strength God
Greatest Weakness Arrogance
Swagger Meter 95%
Black Widow
Greatest Strength Assassin
Greatest Weakness Human
Swagger Meter 75%
Hawkeye
Greatest Strength Marksman
Greatest Weakness Human
Swagger Meter 80%


IR
10.0%
A3
25.0%
A5
10.0%
ES
55.0%


ES
15.0%
IR
10.0%
A5
30.0%
45.0%
A3


ES
5.0%
IR
10.0%
A5
40.0%
45.0%
A3

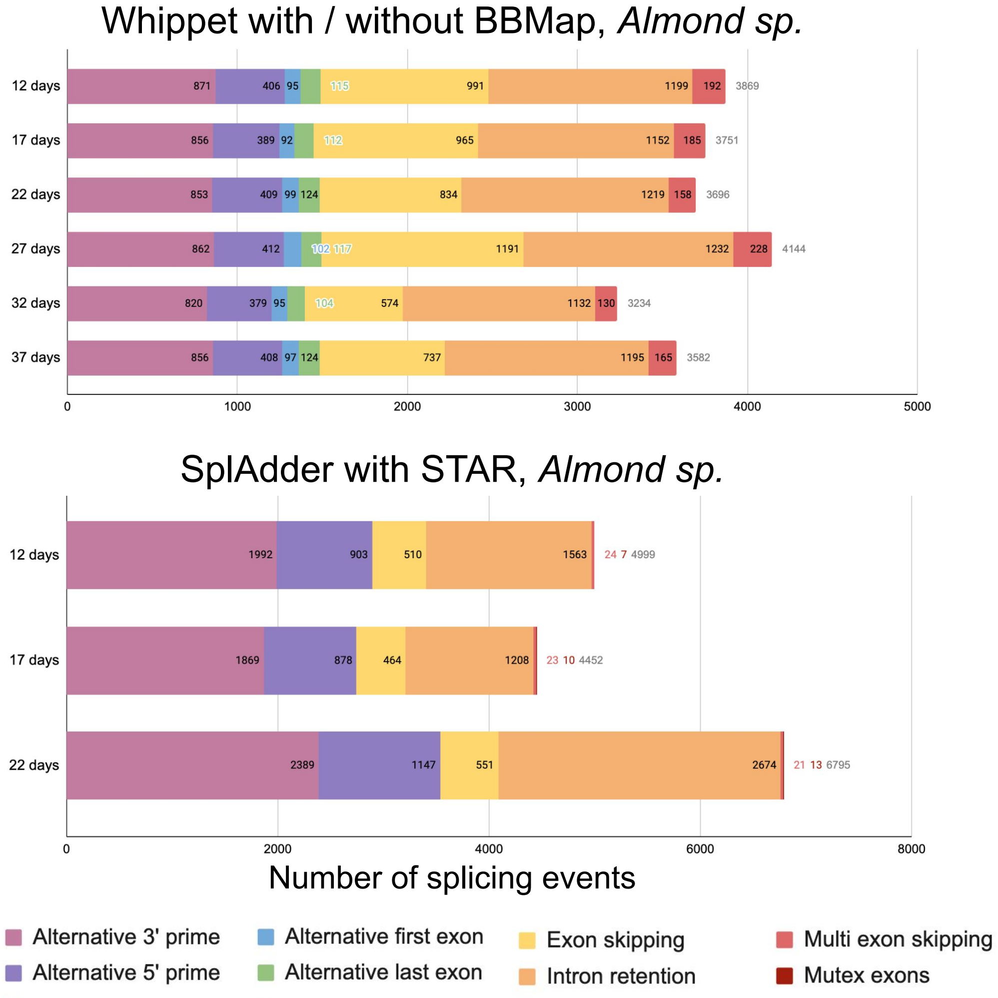


Plan

- Downloading all required information from ENA database
- Quality control
- Alignment
- Data processing
- Alternative splicing programmes
- Visualization of information, received from them
- Analysis of the results


STAR + SPLADDER
BBMAP + WHIPPET

Almond
(Prunus dulcis)


Whippet with / without BBMap, *Almond sp.*

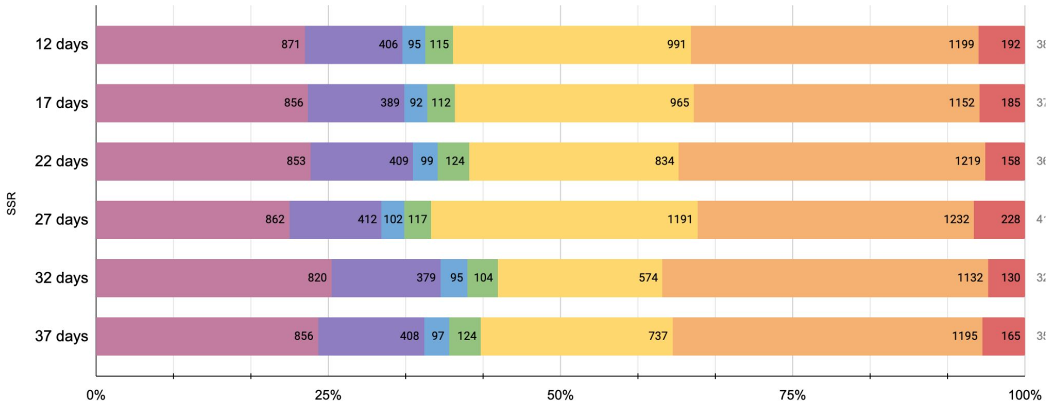
SplAdder with STAR, *Almond sp.*

Comparison of Whippet and SplAdder results for Almond

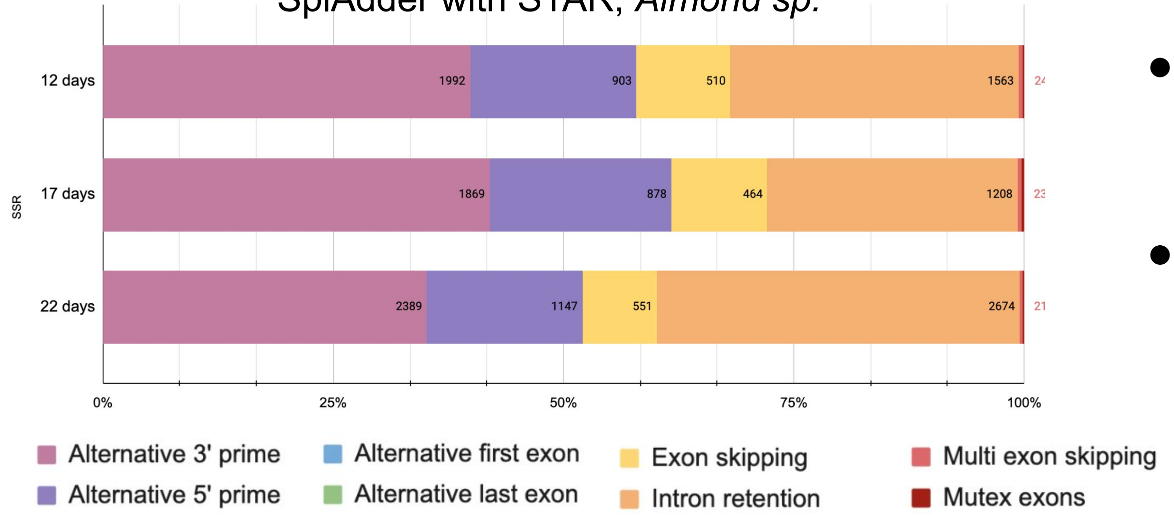
- SplAdder has found more splicing events, however less exon and multi exon skipping
- Whippet can process both fastq/bam files as an input

Comparison of Whippet and SplAdder results for Almond (percents)

Whippet with / without bbmap, *Almond sp.*



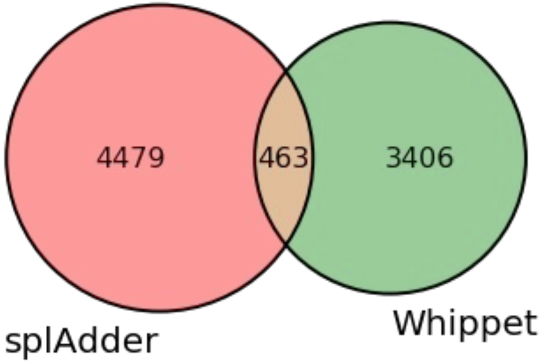
SplAdder with STAR, *Almond sp.*



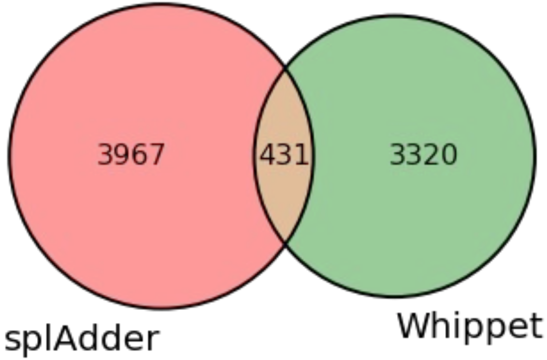
- Both of the programmes are mostly concordant in proportions of IR, A5 and A3
- Proportion of other events differs substantially

STAR + splAdder and bbmap + Whippet ≠ concordance

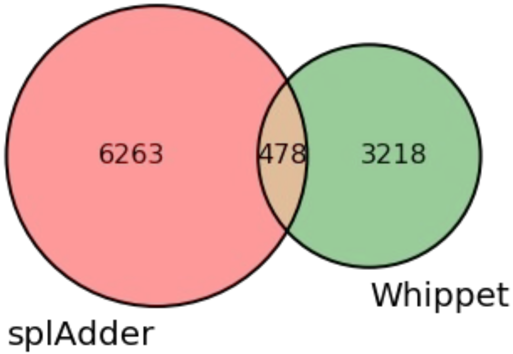
Almond 12 days



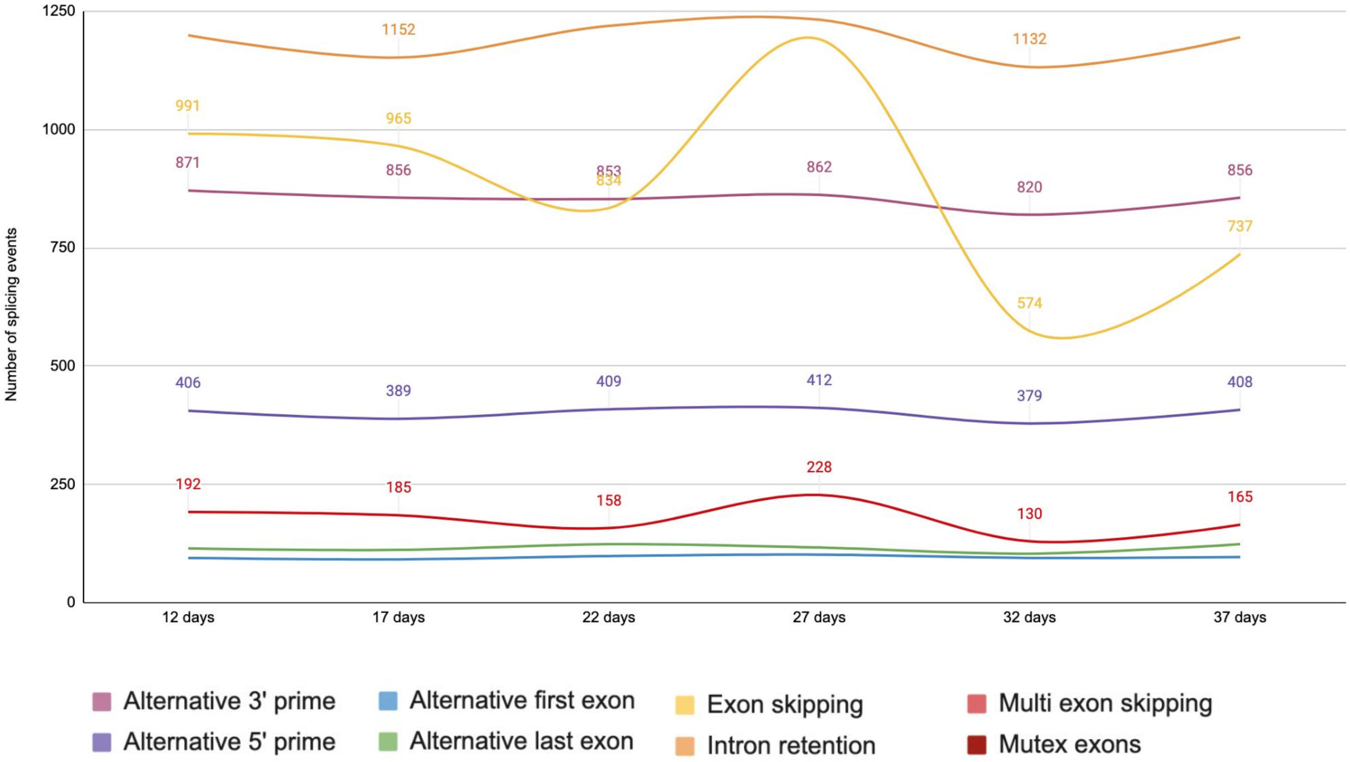
Almond 17 days



Almond 22 days



Distribution of splicing events during the development of *Almond sp.*



Data from Whippet + BBMap

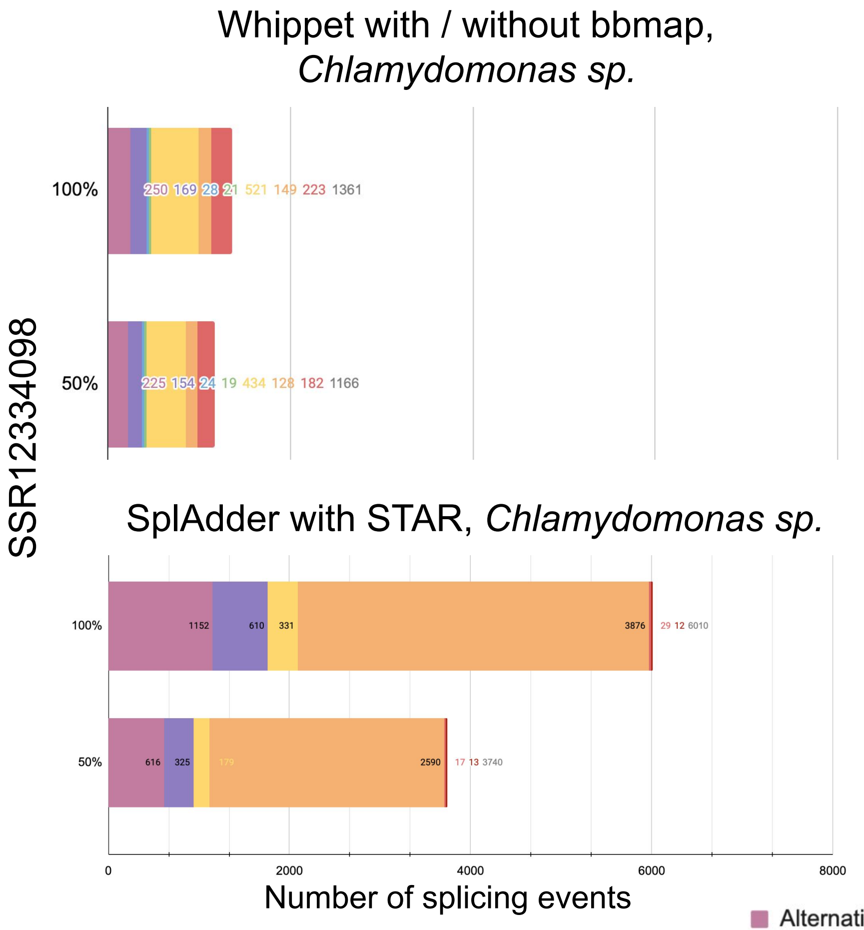
Tissue: almond fruit

- Intron retention remains the most common event through all the development
- We can notice the peak of exon skipping on 27th day

Chlamydomonas
(Chlamydomonas reinhardtii)

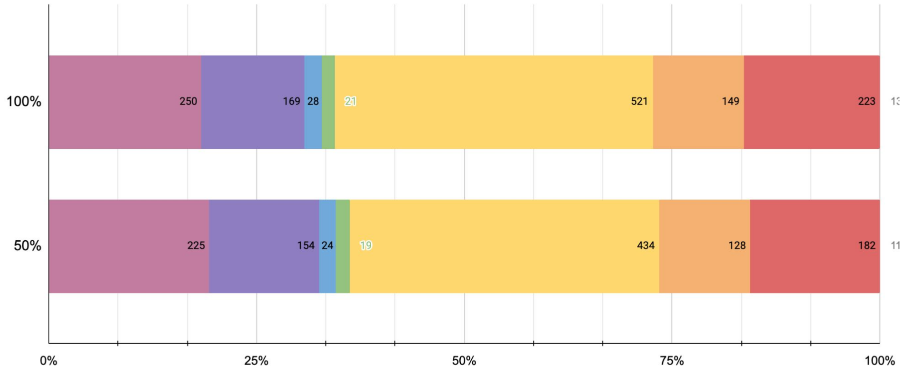
Comparison of Whippet and SplAdder results for Chlamydomonas

- SplAdder has found more splicing events
- Number of reads affects the number of found events: the effect is less dramatic for Whippet

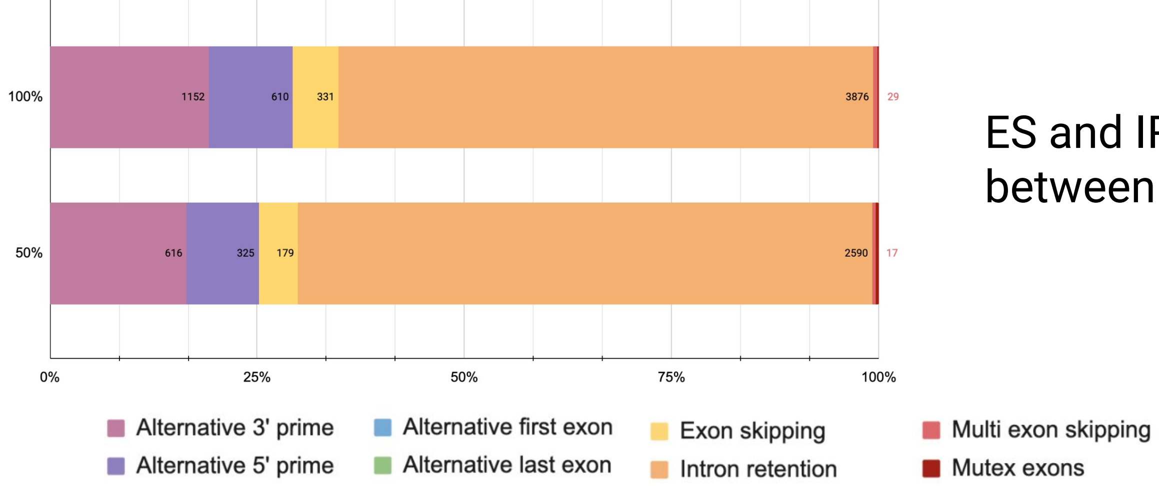


Comparison of Whippet and SplAdder results for Chlamydomonas
(percents)

Whippet with / without bbmap, *Chlamydomonas sp.*

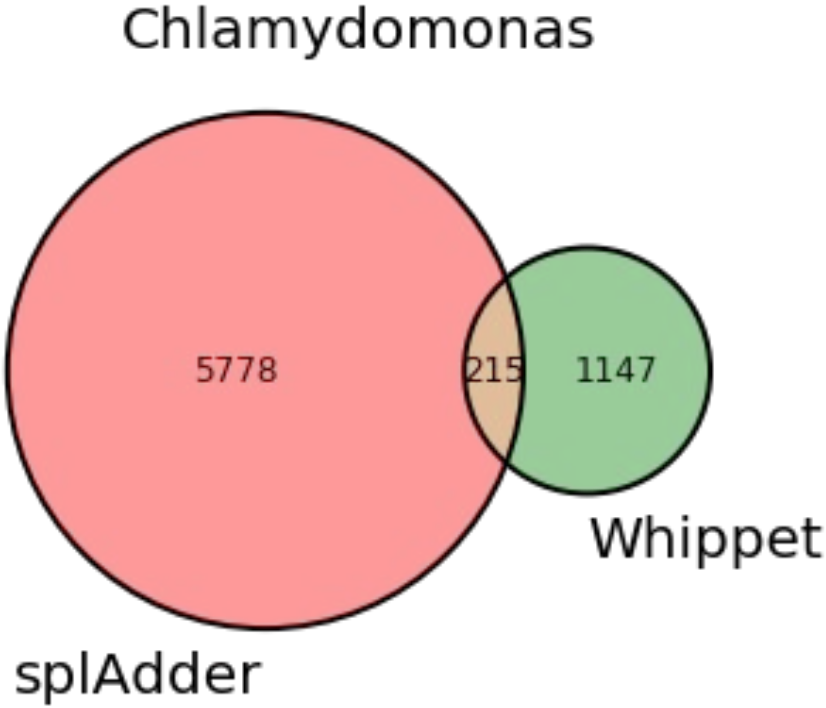


SplAdder with STAR, *Chlamydomonas sp.*



ES and IR varies dramatically between pipelines

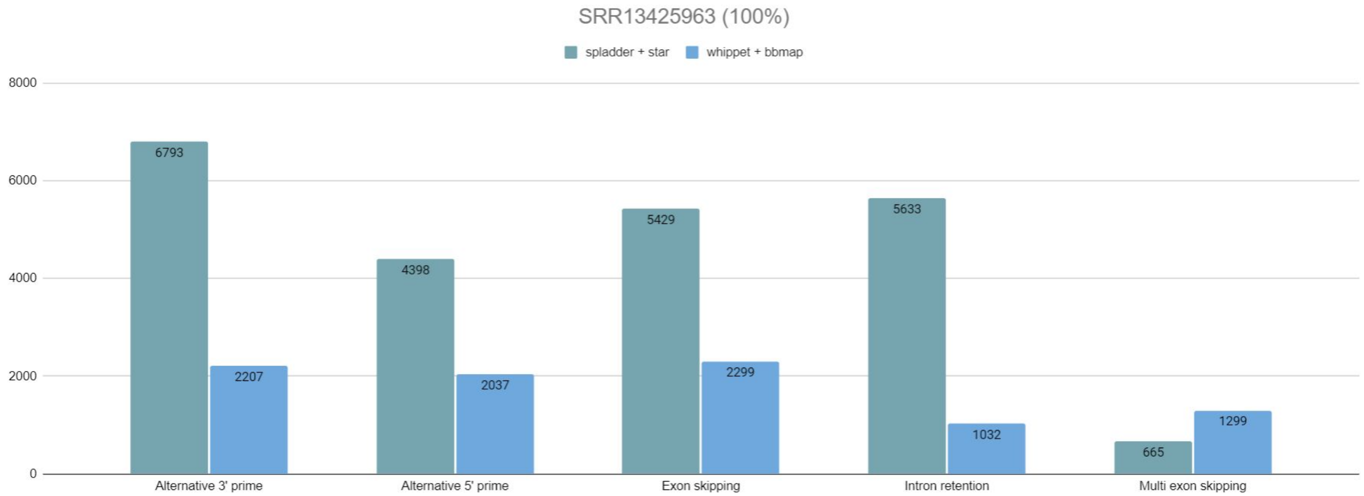
STAR + splAdder and bbmap + Whippet ≠ concordance

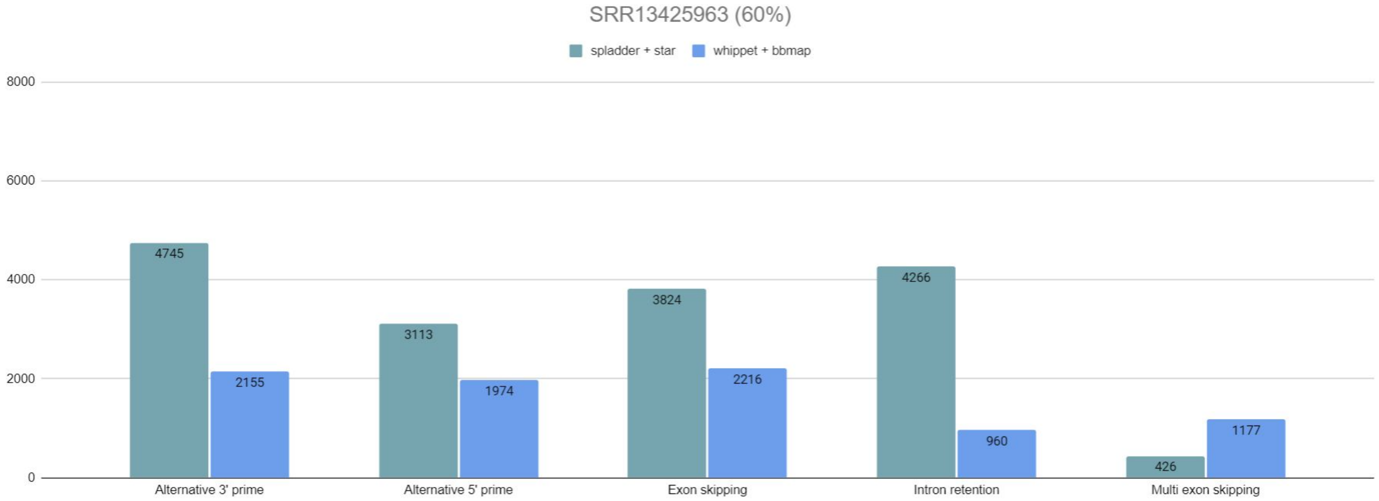


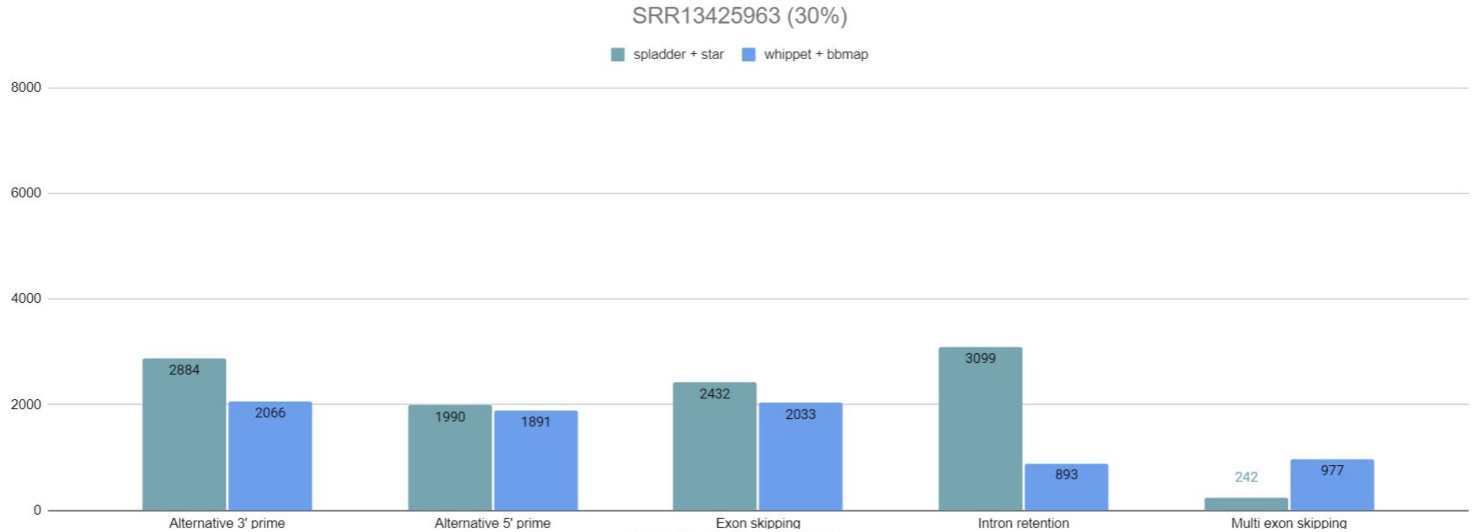
Bee
(*Apis mellifera*)

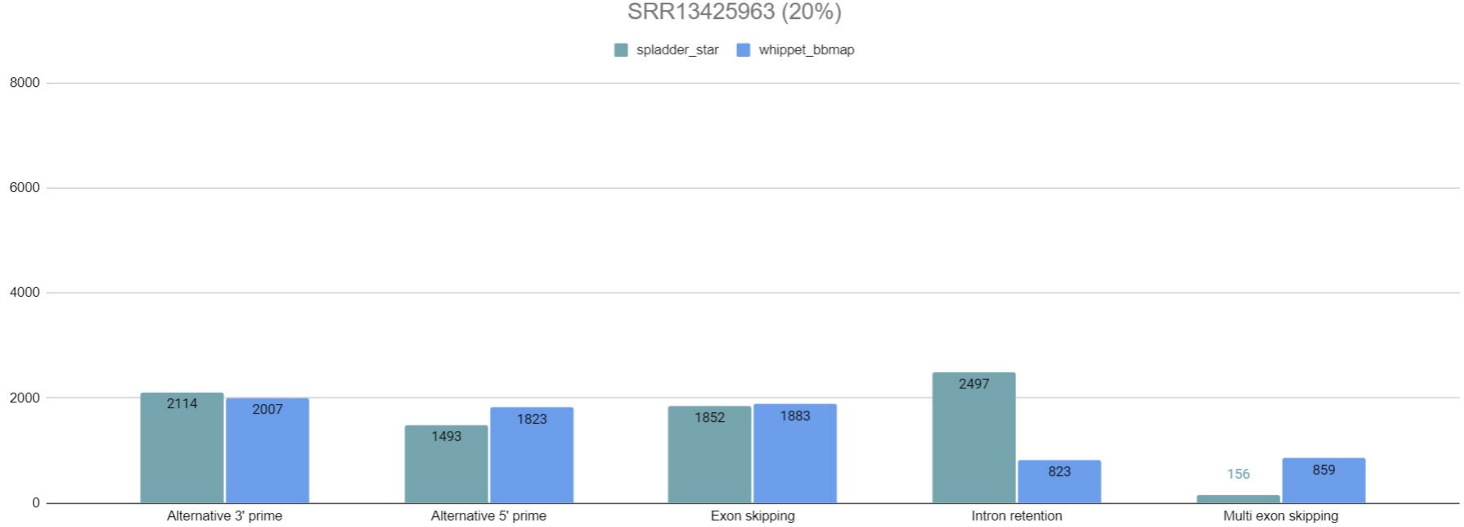
Pipelines comparison

- Downsampling from 30 million reads to 20, 10 and 5 million reads
- Overall number of events was calculated from two pipelines
- Number of reads affects SplAdder but not Whippet results
- For high number of reads SplAdder identify much more events


SRR13425963 (100%)
spladder + star
whippet + bbmap
8000
6000
4000
2000
0
6793
2207
4398
2037
5429
2299
5633
1032
665
1299
Alternative 3' prime
Alternative 5' prime
Exon skipping
Intron retention
Multi exon skipping


SRR13425963 (60%)
spladder + star
whippet + bbmap
8000
6000
4000
2000
0
4745
2155
3113
1974
3824
2216
4266
960
426
1177
Alternative 3' prime
Alternative 5' prime
Exon skipping
Intron retention
Multi exon skipping

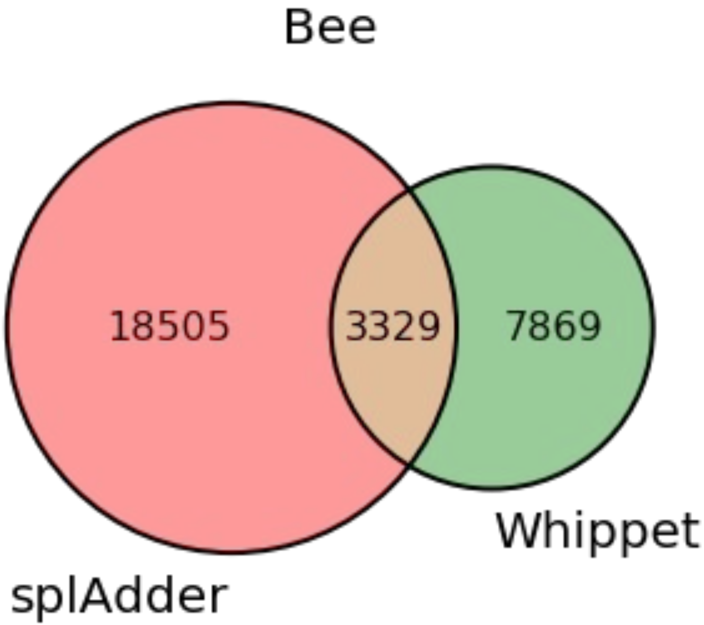

SRR13425963 (30%)


SRR13425963 (20%)

Proportions of all events but IR remain the same in all downsampling experiments

Whippet underestimates IR events

STAR + splAdder and bbmap + Whippet ≠ concordance



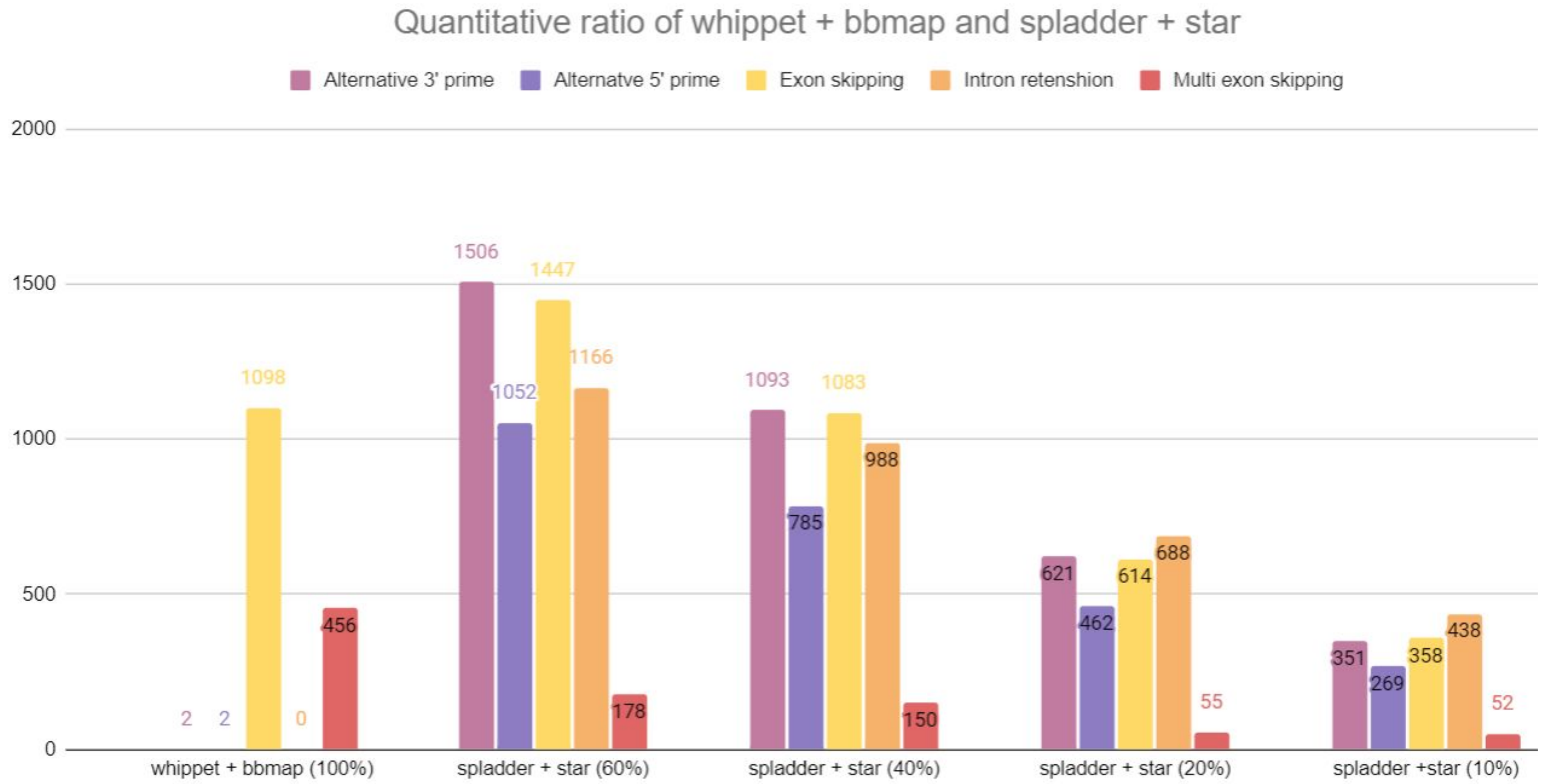
Starlet sea anemone (*Nematostella vectensis*)

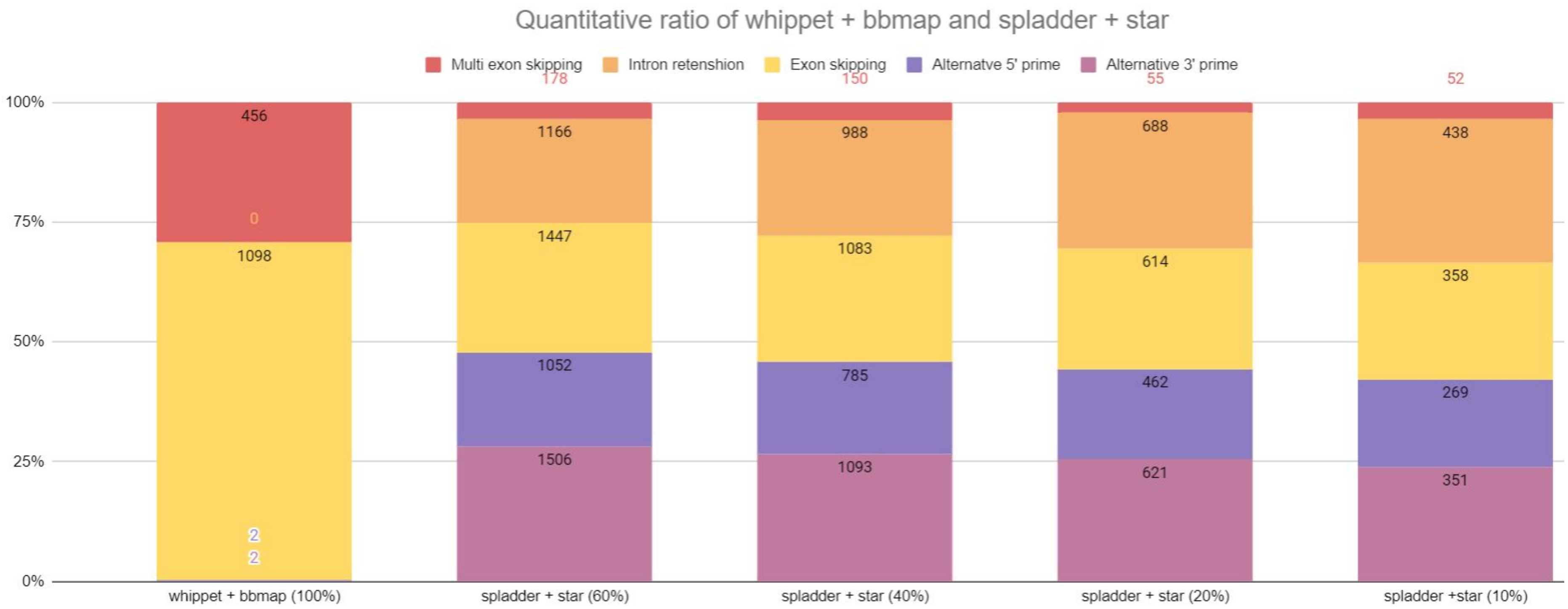
Pipelines comparison

Whippet failed for Nematostella genome - probably, bad annotation

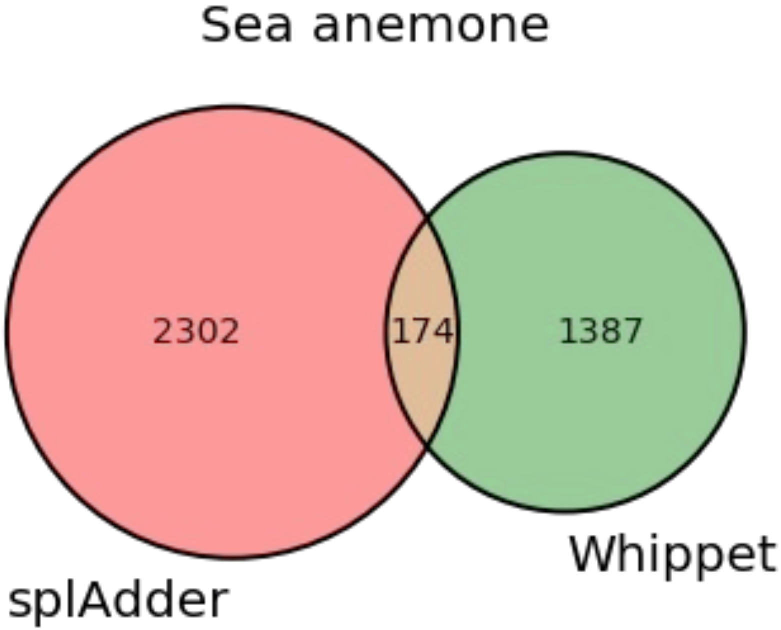
Number of reads affects the results of SplAdder - more reads are better

Proportions of all events remain the same in all downsampling experiments


Quantitative ratio of whippet + bbmap and spladder + star
Alternative 3' prime
Alternatve 5' prime
Exon skipping
Intron retenshion
Multi exon skipping
2000
1500
1000
500
0
2
2
1098
0
456
1506
1052
1447
1166
178
1093
785
1083
988
150
621
462
614
688
55
351
269
358
438
52
whippet + bbmap (100%)
spladder + star (60%)
spladder + star (40%)
spladder + star (20%)
spladder +star (10%)


Quantitative ratio of whippet + bbmap and spladder + star
Multi exon skipping
Intron retenshion
Exon skipping
Alternatve 5' prime
Alternative 3' prime
100%
75%
50%
25%
0%
456
0
1098
2
2
178
1166
1447
1052
1506
150
988
1083
785
1093
55
688
614
462
621
52
438
358
269
351
whippet + bbmap (100%)
spladder + star (60%)
spladder + star (40%)
spladder + star (20%)
spladder +star (10%)

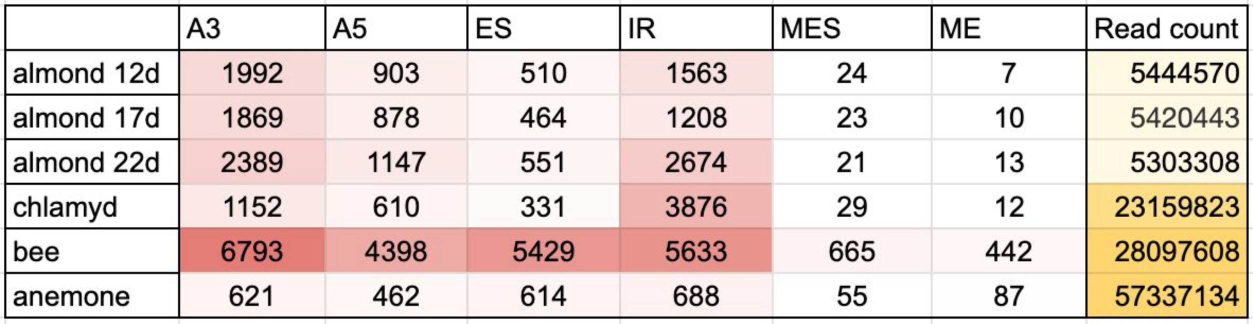
STAR + splAdder and bbmap + Whippet ≠ concordance



All organisms

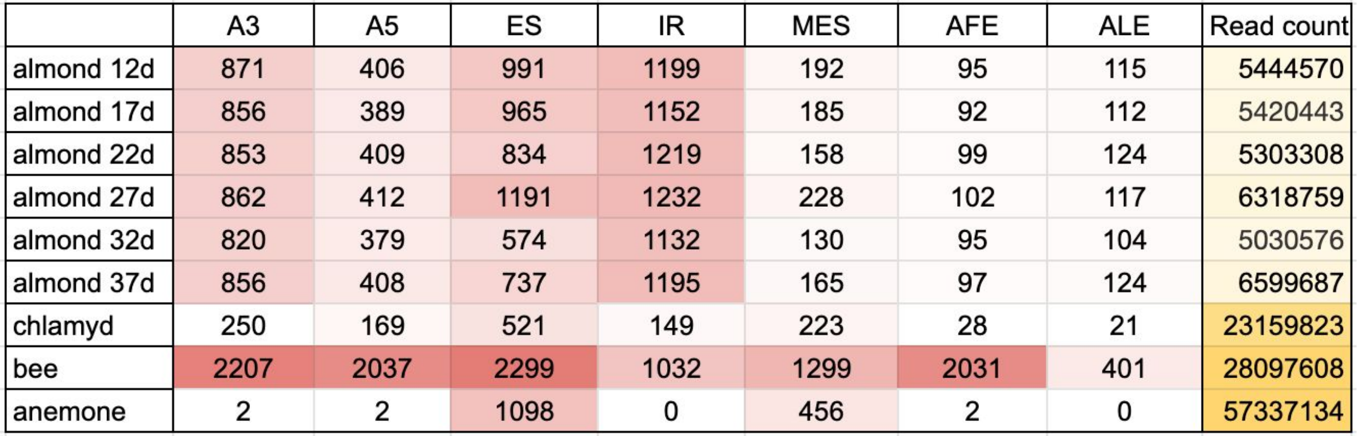
Comparison of results from programs for different organisms

SplAdder



	A3	A5	ES	IR	MES	ME	Read count
almond 12d	1992	903	510	1563	24	7	5444570
almond 17d	1869	878	464	1208	23	10	5420443
almond 22d	2389	1147	551	2674	21	13	5303308
chlamyd	1152	610	331	3876	29	12	23159823
bee	6793	4398	5429	5633	665	442	28097608
anemone	621	462	614	688	55	87	57337134

Whippet

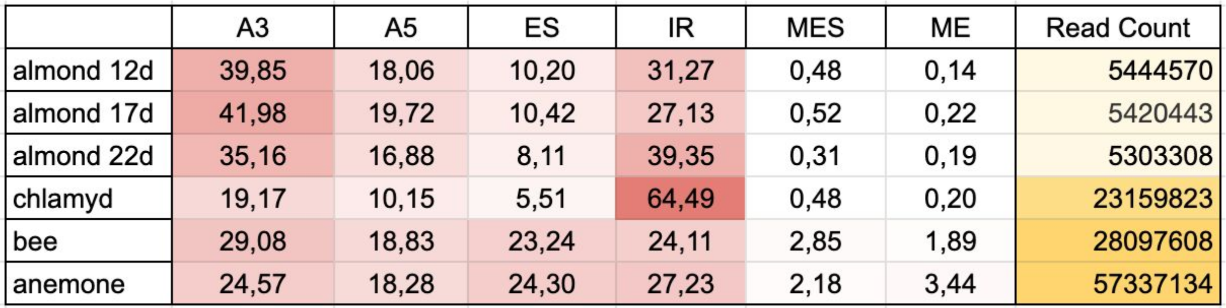


	A3	A5	ES	IR	MES	AFE	ALE	Read count
almond 12d	871	406	991	1199	192	95	115	5444570
almond 17d	856	389	965	1152	185	92	112	5420443
almond 22d	853	409	834	1219	158	99	124	5303308
almond 27d	862	412	1191	1232	228	102	117	6318759
almond 32d	820	379	574	1132	130	95	104	5030576
almond 37d	856	408	737	1195	165	97	124	6599687
chlamyd	250	169	521	149	223	28	21	23159823
bee	2207	2037	2299	1032	1299	2031	401	28097608
anemone	2	2	1098	0	456	2	0	57337134

Comparison of results from programs for different organisms

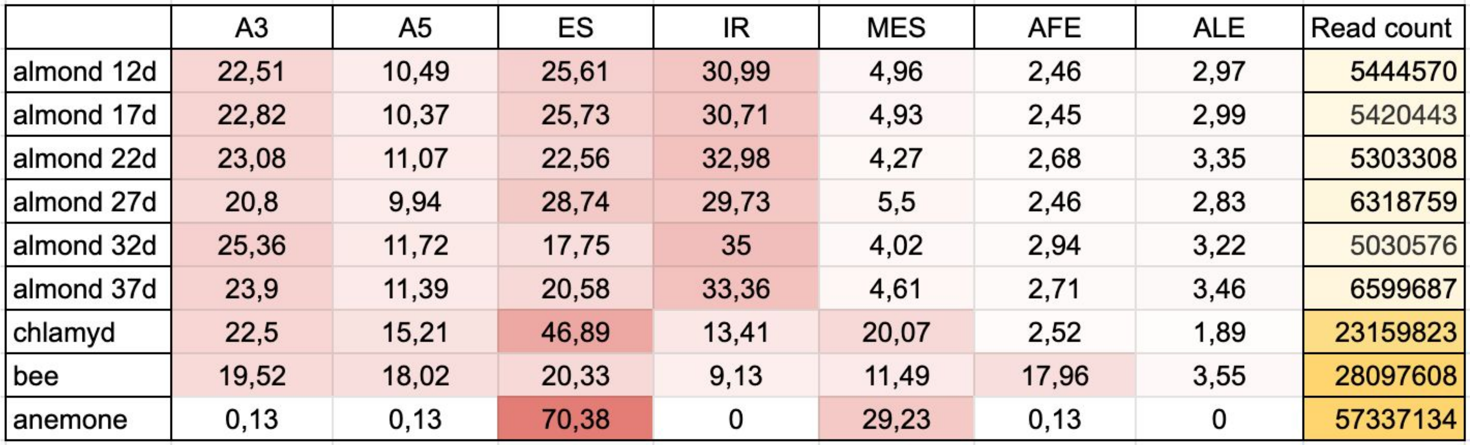
(percents)

SplAdder



	A3	A5	ES	IR	MES	ME	Read Count
almond 12d	39,85	18,06	10,20	31,27	0,48	0,14	5444570
almond 17d	41,98	19,72	10,42	27,13	0,52	0,22	5420443
almond 22d	35,16	16,88	8,11	39,35	0,31	0,19	5303308
chlamyd	19,17	10,15	5,51	64,49	0,48	0,20	23159823
bee	29,08	18,83	23,24	24,11	2,85	1,89	28097608
anemone	24,57	18,28	24,30	27,23	2,18	3,44	57337134

Whippet



	A3	A5	ES	IR	MES	AFE	ALE	Read count
almond 12d	22,51	10,49	25,61	30,99	4,96	2,46	2,97	5444570
almond 17d	22,82	10,37	25,73	30,71	4,93	2,45	2,99	5420443
almond 22d	23,08	11,07	22,56	32,98	4,27	2,68	3,35	5303308
almond 27d	20,8	9,94	28,74	29,73	5,5	2,46	2,83	6318759
almond 32d	25,36	11,72	17,75	35	4,02	2,94	3,22	5030576
almond 37d	23,9	11,39	20,58	33,36	4,61	2,71	3,46	6599687
chlamyd	22,5	15,21	46,89	13,41	20,07	2,52	1,89	23159823
bee	19,52	18,02	20,33	9,13	11,49	17,96	3,55	28097608
anemone	0,13	0,13	70,38	0	29,23	0,13	0	57337134

Conclusion

Pipeline for alternative splicing analysis is critical for the results cause of:

- Number of reads
- Annotation quality

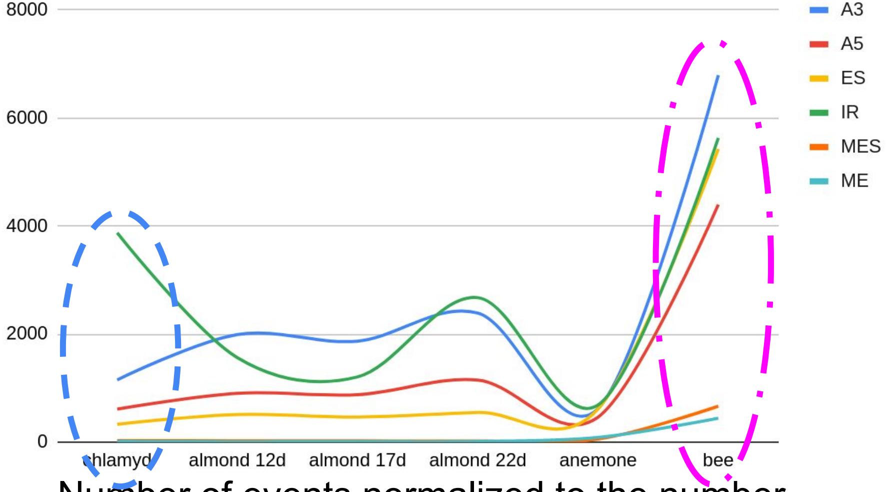
Whippet mostly fails for poorly characterized genomes (chlamydomonas and sea anemone)

Results of pipelines differ significantly: which to believe -> next SMTB project

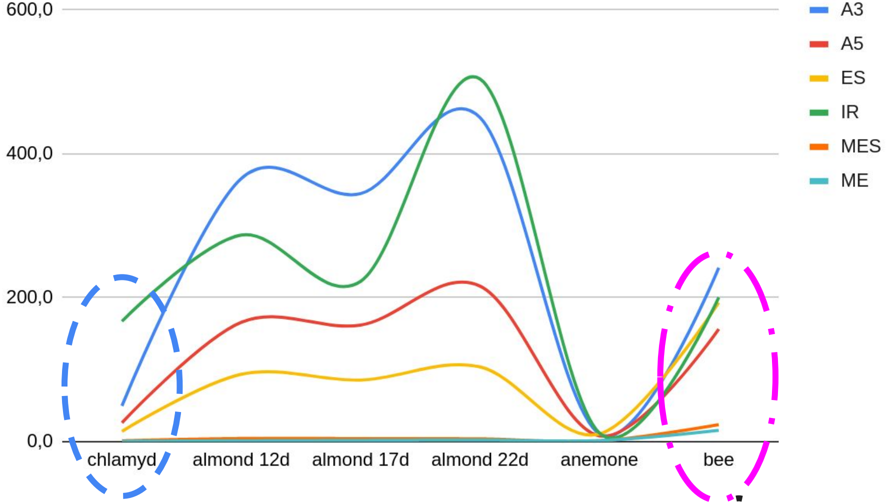
Alternative splicing events proportions and number changes:

- from organism to organism
- during the development

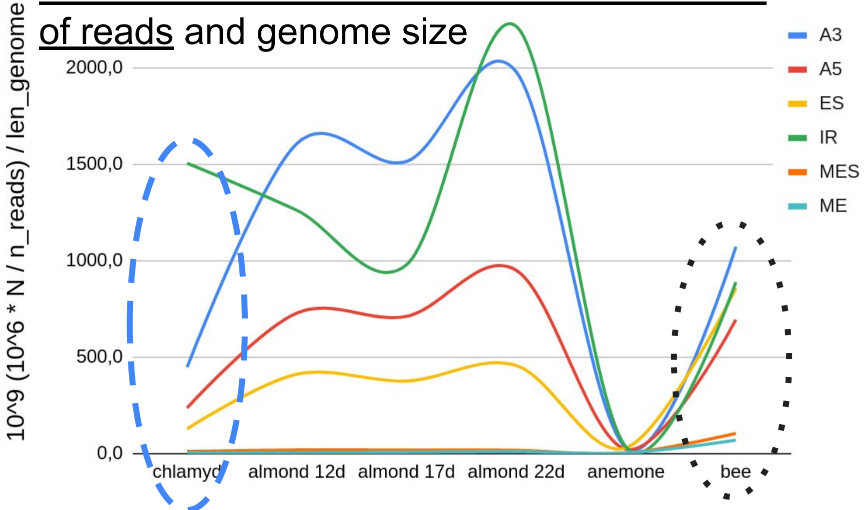
Number of events



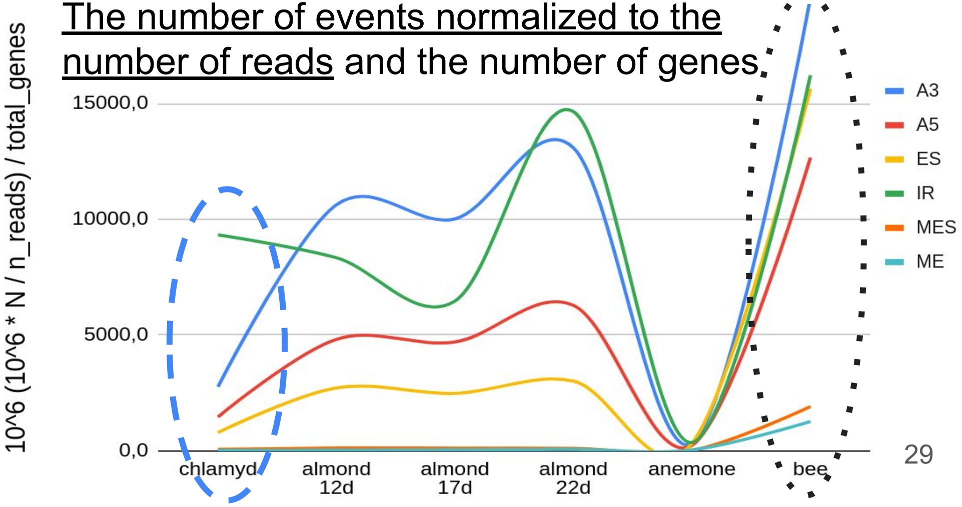
Number of events normalized to the number of reads



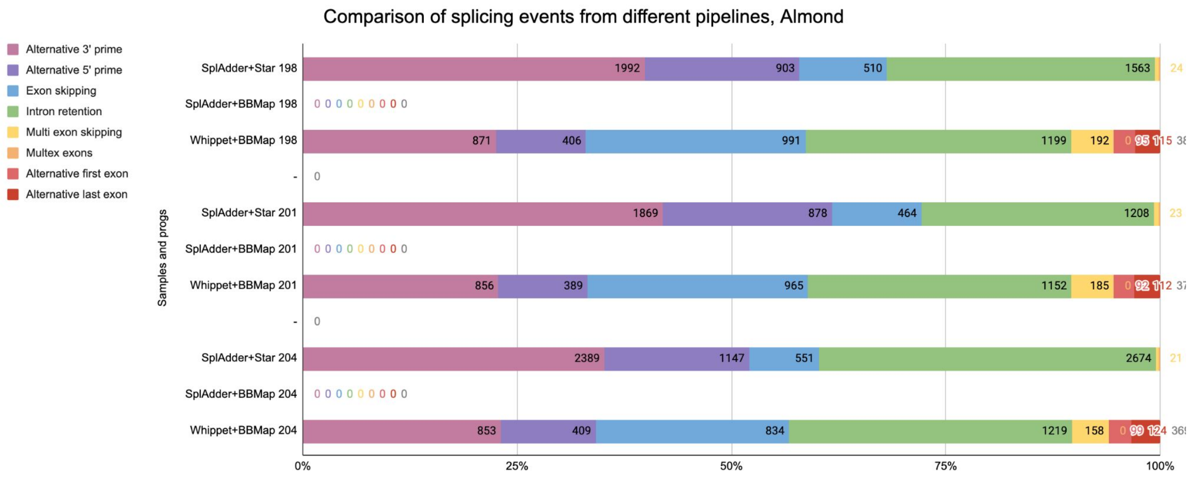
Number of events normalized to the number of reads and genome size

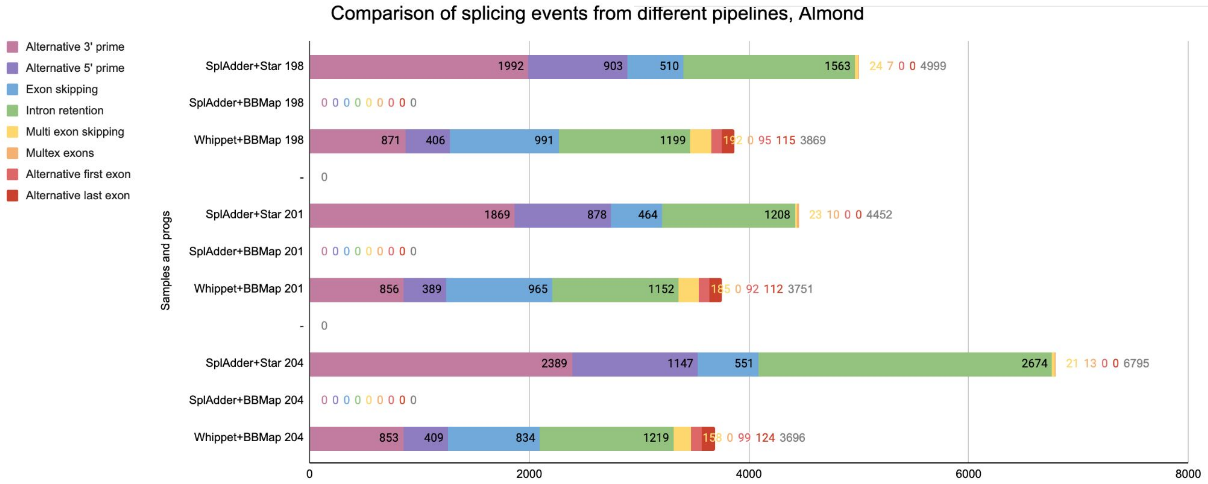


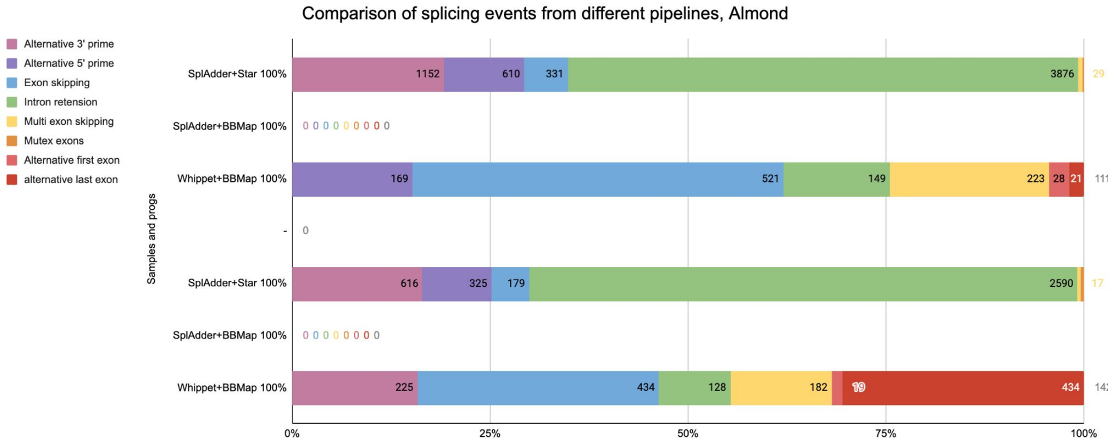
The number of events normalized to the number of reads and the number of genes.

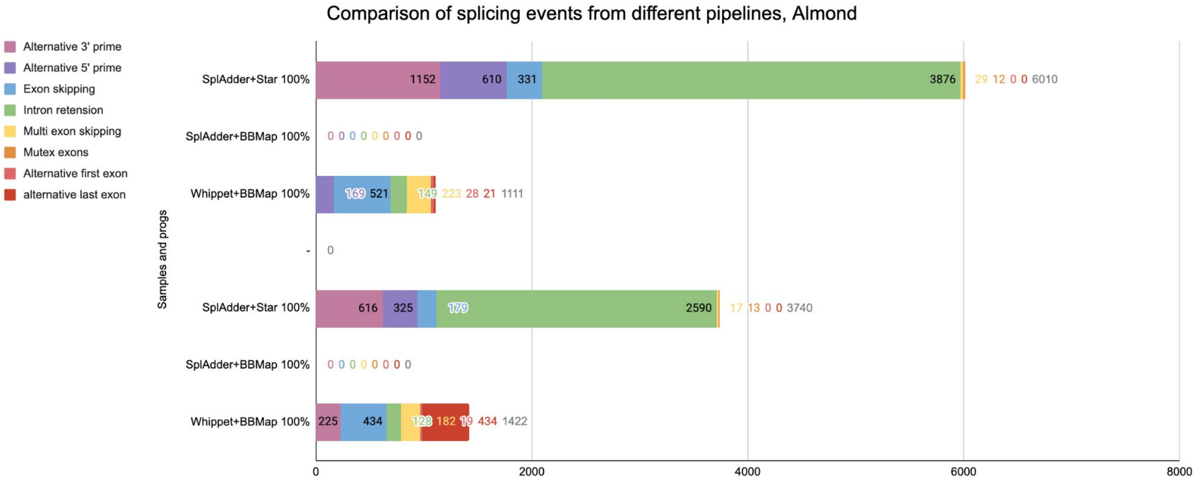


Name	Read count	Golden Path Length	Total genes	Coding genes	Non coding genes
Chlamydomonas reinhardtii	23’159’823	111’098’438	17’943	17743	200
Prunus dulcis	5’420’443	227’498’357	34’421	27966	6455
	5’303’308	227’498’357	34’421	27966	6455
	5’420’443	227’498’357	34’421	27966	6455
Nematostella vectensis	57’337’134	356’613’585	32’688	24773	7915
Apis mellifera	28’097’608	225’250’884	12’356	9944	2412


Comparison of splicing events from different pipelines, Almond


Comparison of splicing events from different pipelines, Almond


Comparison of splicing events from different pipelines, Almond
Alternative 3' prime
Alternative 5' prime
Exon skipping
Intron retension
Multi exon skipping
Mutex exons
Alternative first exon
alternative last exon
Samples and progs
SplAdder+Star 100%
1152 610 331 3876 29
SplAdder+BBMap 100%
0 0 0 0 0 0 0 0 0
Whippet+BBMap 100%
169 521 149 223 28 21 111
-
0
SplAdder+Star 100%
616 325 179 2590 17
SplAdder+BBMap 100%
0 0 0 0 0 0 0 0
Whippet+BBMap 100%
225 434 128 182 19 434 14
0%
25%
50%
75%
100%


Comparison of splicing events from different pipelines, Almond
Alternative 3' prime
Alternative 5' prime
Exon skipping
Intron retension
Multi exon skipping
Mutex exons
Alternative first exon
alternative last exon
Samples and progs
SplAdder+Star 100%
1152 610 331 3876 29 12 0 0 6010
SplAdder+BBMap 100%
0 0 0 0 0 0 0 0 0
Whippet+BBMap 100%
169 521 149 223 28 21 1111
-
0
SplAdder+Star 100%
616 325 179 2590 17 13 0 0 3740
SplAdder+BBMap 100%
0 0 0 0 0 0 0 0
Whippet+BBMap 100%
225 434 128 182 19 434 1422
0
2000
4000
6000
8000