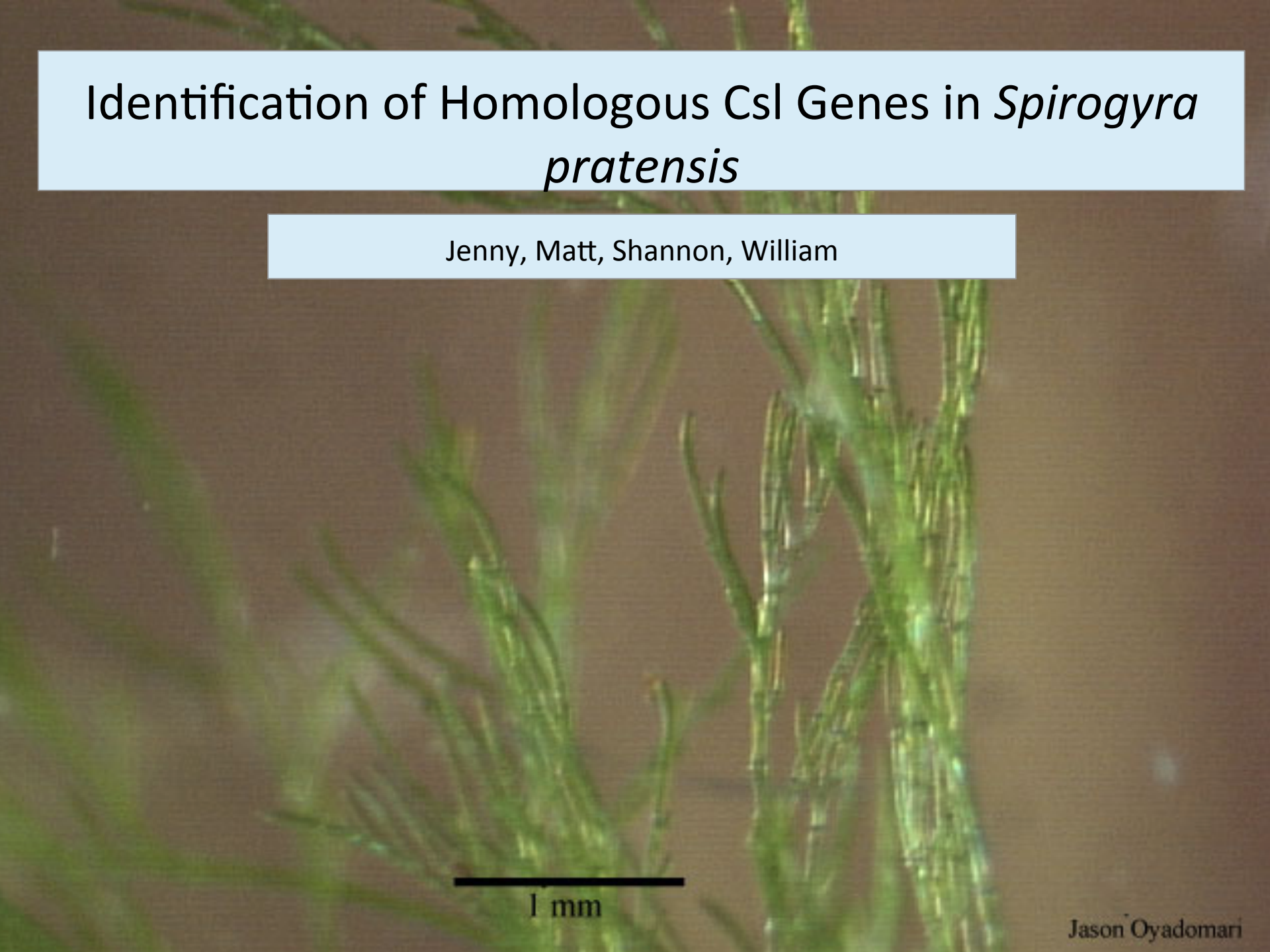


Identification of Homologous Csl Genes in *Spirogyra pratensis*

Jenny, Matt, Shannon, William



1 mm

Cellulose

Cellulose

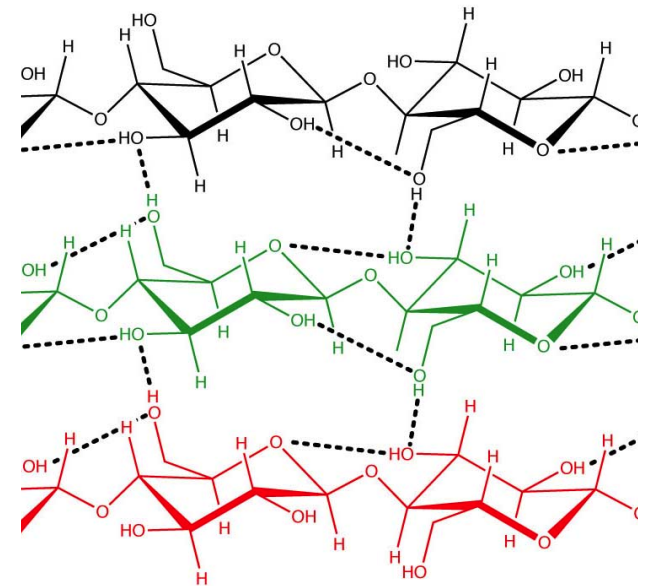
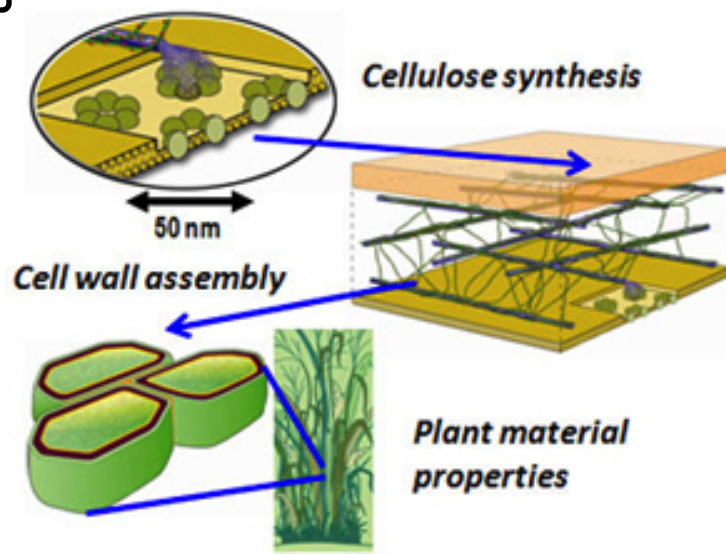
Background

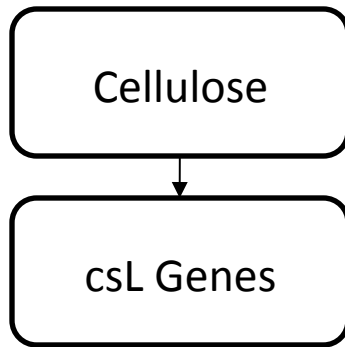
Analysis

Results

Cellulose is a polysaccharide chain of carbohydrates

- Structurally important cell walls, plants, algae, oomycetes.
- Cellulose accounts for 50-70% (dry weight) in biomass
- Plant and Microbial Cellulose





Cellulose like Genes (Csl genes)

Csl genes encode cellulose synthase enzymes

➤ Nine Csl subfamilies

Csl Subfamilies	
➤ CesA family	➤ CslE
➤ CslA	➤ CslF
➤ CslB	➤ CslG
➤ CslC	➤ CslH
➤ CslD	

BMC Plant Biology 2009 Sequences

Cellulose

csL Genes

csL Database

Background

Analysis

Results

A database of identified Csl genes from 17 species

Table 1: Plant and algal genomes used in the present study

Index	Abbr.	Clade	Species	Genome Published/Released	Csl Published?
1	Tp	Diatom	<i>Thalassiosira pseudonana</i>	[25]	N
2	Pht	Diatom	<i>Phaeodactylum tricornutum</i>	JGI	N
3	Pa	brown tide algae	<i>Aureococcus anophagefferens</i>	JGI	N
4	Cm	red algae	<i>Cyanidioschyzon merolae</i>	[26]	N
5	Mpc	green algae	<i>Micromonas pusilla CCMP1545</i>	[27]	N
6	Mpr	green algae	<i>Micromonas strain RCC299</i>	[27]	N
7	Ol	green algae	<i>Ostreococcus lucimarinus</i>	[28]	N
8	Ot	green algae	<i>Ostreococcus tauri</i>	[29]	N
9	Cr	green algae	<i>Chlamydomonas reinhardtii</i>	[30]	N
10	Vc	green algae	<i>Volvox carteri f. nagariensis</i>	JGI	N
11	Pp	moss	<i>Physcomitrella patens ssp patens</i>	[31]	[6]
12	Sm	spike moss	<i>Selaginella moellendorffii</i>	JGI	N [32]
13	Pt	dicot	<i>Populus trichocarpa</i>	[33]	[5]
14	At	dicot	<i>Arabidopsis thaliana</i>	[34]	[2]
15	Vv	dicot	<i>Vitis vinifera</i>	[35]	N
16	Os	monocot	<i>Oryza sativa</i>	[36,37]	[3]
17	Sb	monocot	<i>Sorghum bicolor</i>	[38]	N

Cellulose



csL Genes



csL Database



Charophyta

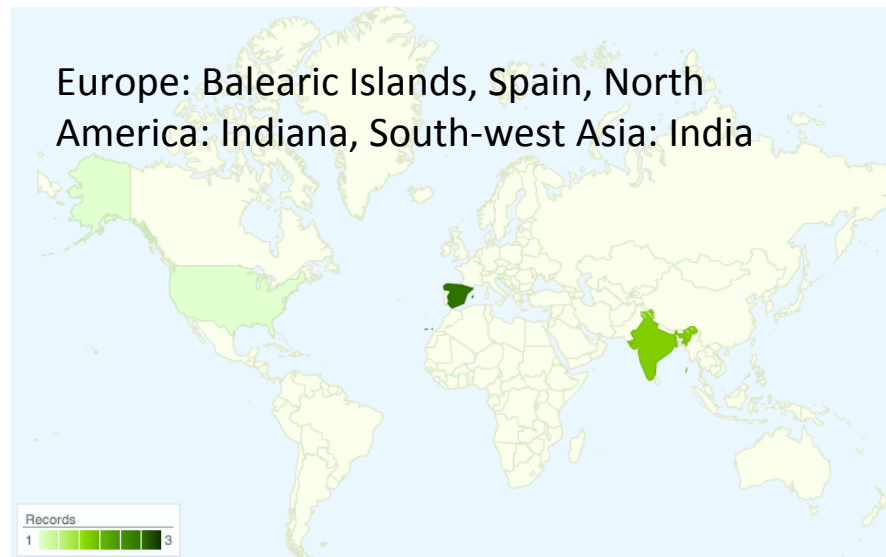
Spirogyra pratensis

A Fresh water algae (charophyta)

In 2010, Ruth E. Timme¹ and Charles F. Delwiche published SRA data sets of *Spirogyra pratensis*.

We wanted to know if there were and Csl homologs in the previously published database.

Distribution by country



http://www.algaebase.org/search/species/detail/?species_id=59533

Background

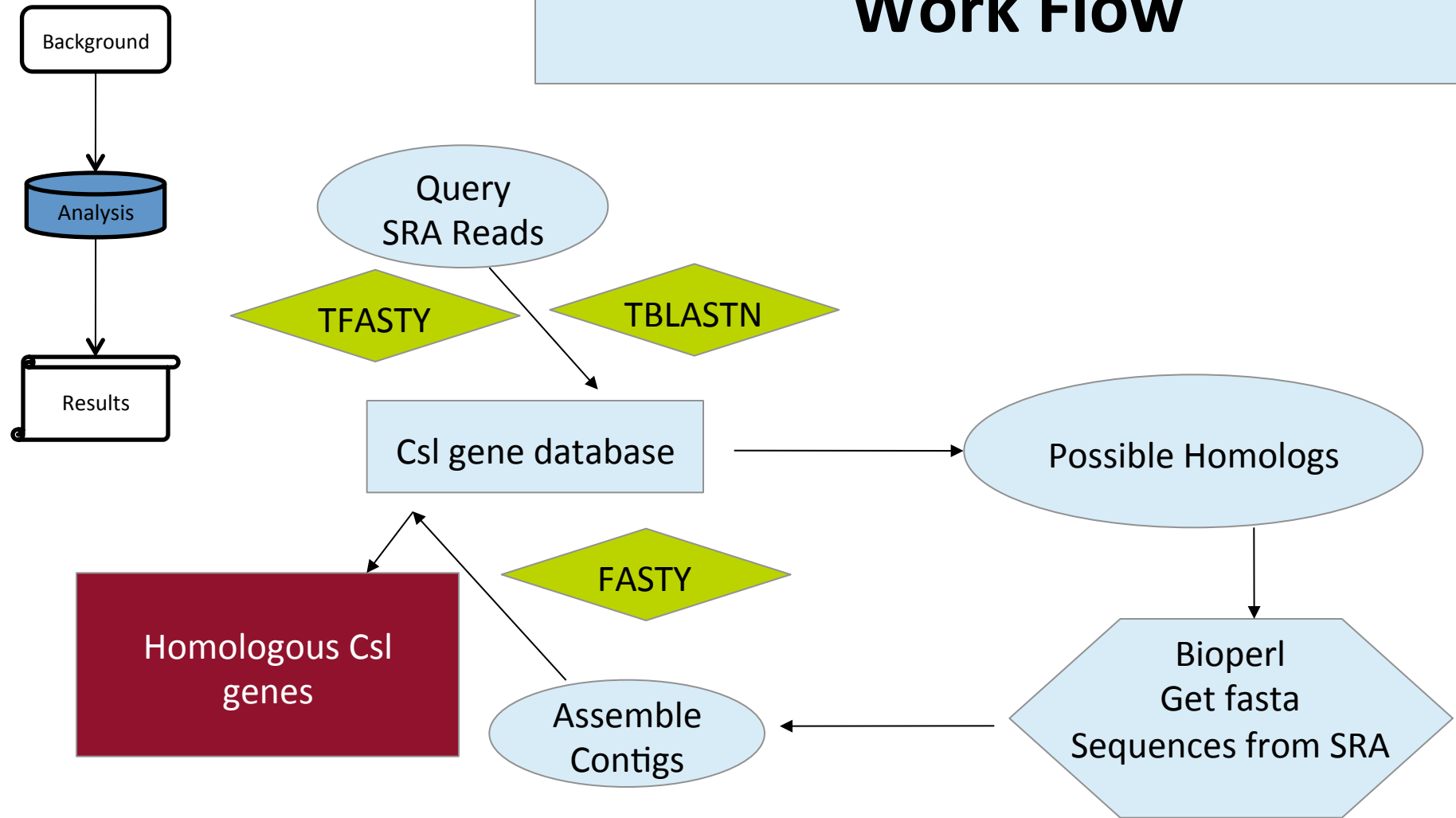


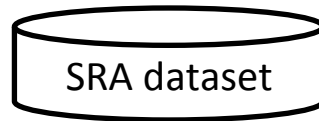
Analysis



Results

Work Flow





Spirogyra pratensis datasets were downloaded from NCBI.

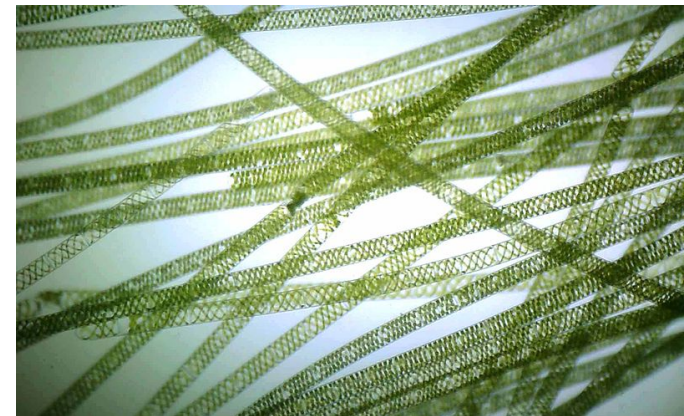
This sra library is the result of genomic work performed at UTEX as part of combination sanger 454 study of two charophyte species:

Spirogyra pratensis
and
Coleochaete orbicularis.

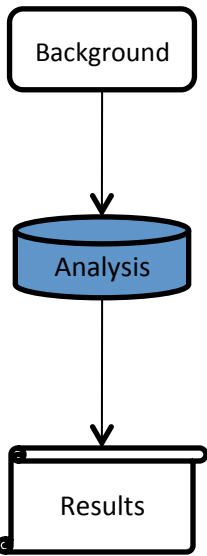
The post clustering *S. Pratensis* dataset contains over 12,000 unique transcripts

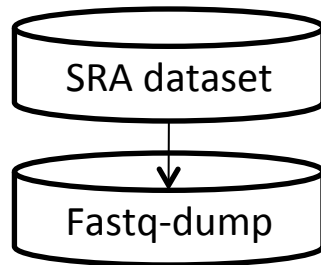


Pond Scum Profile: *Spirogyra* is cloud forming with long filaments, generally unbranched.



Spirogyra pratensis image borrowed from Charles F. Delwiche lab at the University of Maryland



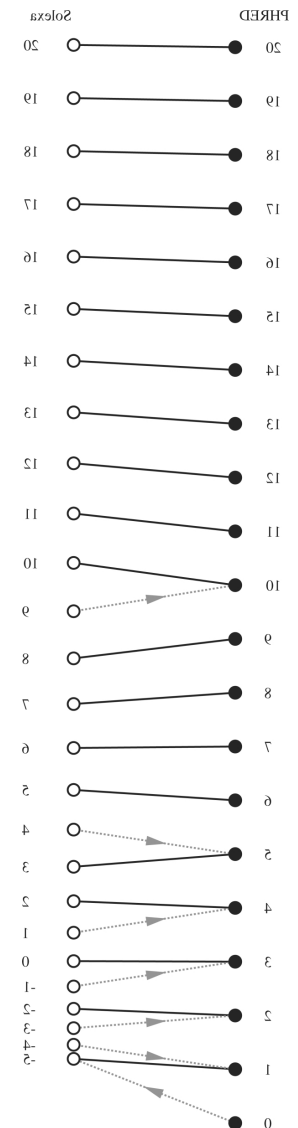


Using the Fastq-dump from the SRA toolkit, we extract the sequence information from the SRA data along with the quality measure of the read.

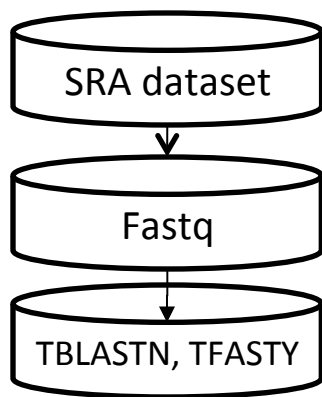
```

@SRR014849.1 EIXKN4201CFU84 length=93
GGGGGGGGGGGGGGGGGCTTTTTTGTGGGAACCGAAAGG
GTTTGAATTTCAAACCTTTTCGGTTTCCAACCTTCCAA
AGCAATGCCAATA
+SRR014849.1 EIXKN4201CFU84 length=93
3+&$#"7F@71,'";C?,B;?6B;:EA1EA
1EA5'9B?:#9EA0D@2EA5':>5?:%A;A8A;?9B;D@
/=<?7=9<2A8==
  
```

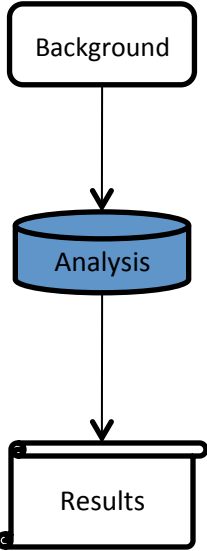
*@title and optional description
sequence line(s)
+optional repeat of title line
quality line(s)*



Visual representation of the mapping between PHRED and Solexa quality scores. Vertical layout represents the probability of error on a log scale, therefore the PHRED points are equally spaced (black circles on left), while the Solexa points are not (white circles on right). Solid black lines are reciprocal mappings between scores, and grey arrows are lossy mappings. Near the top of the figure, the black lines are almost horizontal because the two scores are almost equal. The straightforward mapping of higher scores is omitted due to space.



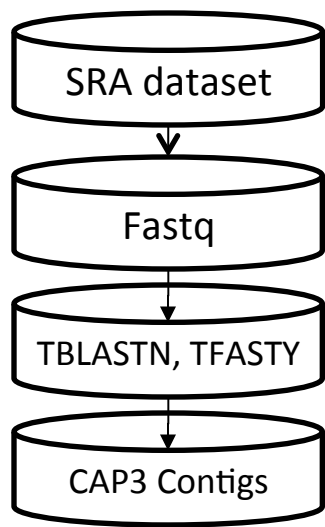
Identification of possible Homologs



The SRA reads were translated into protein sequences and used as a query against the published Csl database.

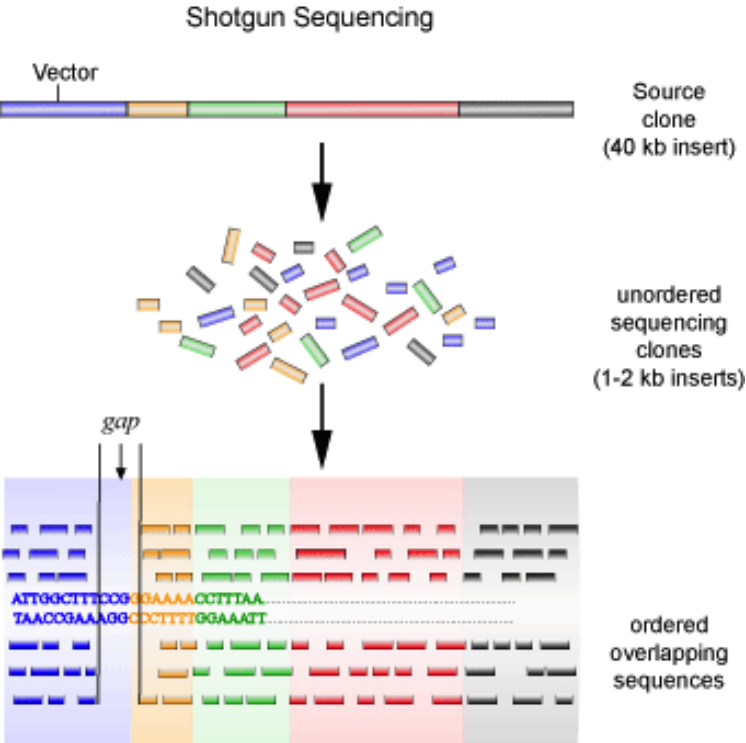
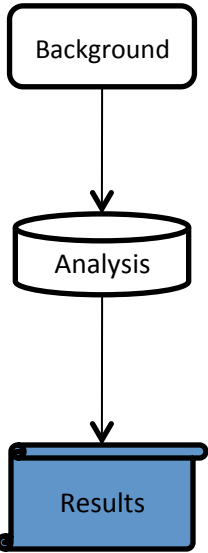
Homologous hits	
TBLASTN	TFASTY
91126	94904

AT2G21770.1	AT2G21770.1	cesA	SRR036731.1335640	87.84	74	9	0	737	830	17	238	9e-41	143
AT2G21770.1	AT2G21770.1	cesA	SRR036731.352599	85.00	80	11	1	787	865	249	10	1e-40	145
AT2G21770.1	AT2G21770.1	cesA	SRR036731.125813	80.25	81	15	1	816	895	257	15	2e-40	144
AT2G21770.1	AT2G21770.1	cesA	SRR036731.45630	85.71	77	11	0	481	557	15	245	3e-40	144
AT2G21770.1	AT2G21770.1	cesA	SRR036731.313154	80.25	81	15	1	815	894	257	15	3e-40	144
AT2G21770.1	AT2G21770.1	cesA	SRR036731.257160	86.49	74	10	0	488	561	240	19	5e-40	143
AT2G21770.1	AT2G21770.1	cesA	SRR036731.16555	81.33	75	14	0	515	589	17	241	7e-40	143
AT2G21770.1	AT2G21770.1	cesA	SRR036731.33011	81.25	80	14	1	809	887	14	253	1e-39	142
AT2G21770.1	AT2G21770.1	cesA	SRR036731.94426	77.38	84	18	1	781	863	12	263	1e-39	142
AT2G21770.1	AT2G21770.1	cesA	SRR036731.164760	76.00	75	18	0	582	656	24	248	1e-39	142
AT2G21770.1	AT2G21770.1	cesA	SRR036731.48220	80.25	81	16	0	417	497	18	260	3e-39	141
AT2G21770.1	AT2G21770.1	cesA	SRR036731.118122	75.00	76	19	0	555	630	1	228	8e-39	140
AT2G21770.1	AT2G21770.1	cesA	SRR036731.148207	76.83	82	18	1	818	898	16	261	8e-39	140
AT2G21770.1	AT2G21770.1	cesA	SRR036731.32898	77.11	83	19	1	739	821	258	16	2e-38	139
AT2G21770.1	AT2G21770.1	cesA	SRR036731.103397	88.57	70	8	0	769	838	16	225	6e-38	137



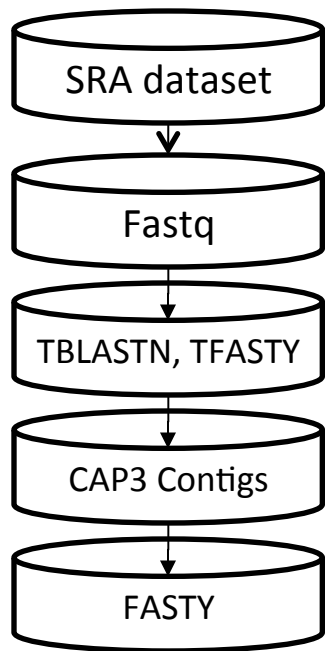
Contiq Construction

CAP3 “Shotgun” assembly



74 assembled contigs.

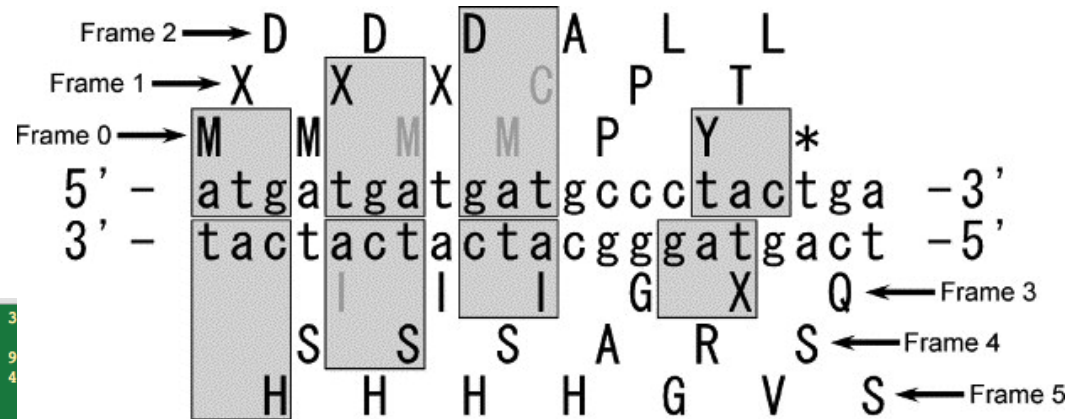
CAP3:
Xiaoqiu Huang and Anup Madan, 1999



Homolog Confirmatory Step

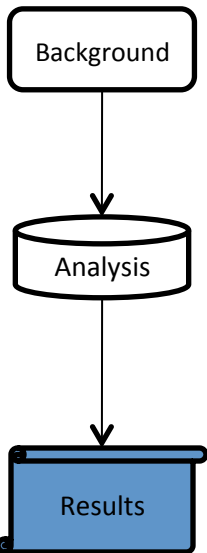
FASTY- DNA protein to a Protein database

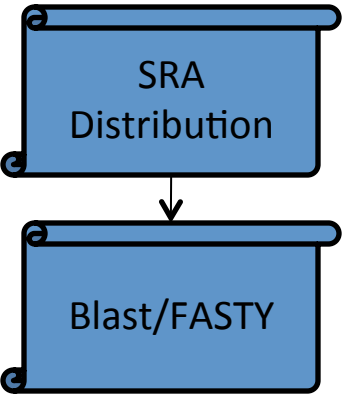
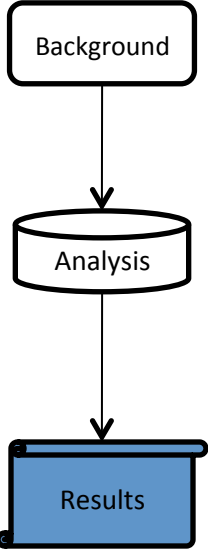
- Using 3 frame shifts
- E-value cut off 1e-5
- Best match was taken



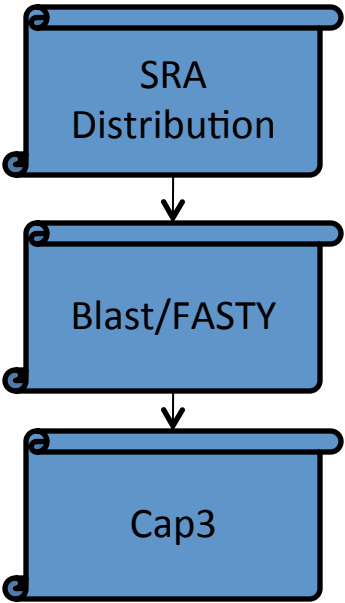
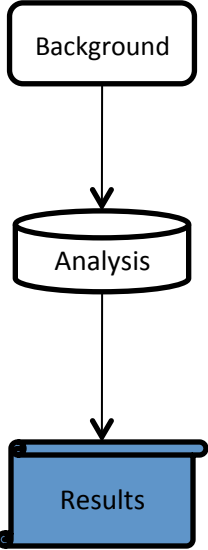
```

18 sm_18364|estExt_Genewise1Plus.C_490155|cesA 3
23 os_42916|LOC_Os07g36630.1|cs1F 33 1.9e-07
25 sb_5965|estExt_Genewise1.C_chr_11372|cs1C 9
29 pp_2722|estExt_Genewise1.C_80120|cesA 108 4
31 os_5702|LOC_Os01g54620.1|cesA 92 7.7e-36
33 vv_19270|GSVIVP00025577001|cesA 135 2.6e-19
34 pp_8706|estExt_fggenesh1_pm.C_300004|cesA 139 6.1e-63
35 os_53309|LOC_Os09g39920.1|cs1A 79 1.9e-08
37 pp_4876|e_gw1.29.2.1|cesA 270 3.8e-63
41 pt_30735|eugene3.00160483|cesA 39 1.8e-12
45 pp_5273|estExt_gwp_gw1.C_1890065|cesA 117 1.1e-26
50 vv_22004|GSVIVP00028800001|cesA 136 1.1e-49
57 os_53309|LOC_Os09g39920.1|cs1A 78 1.6e-10
60 os_53309|LOC_Os09g39920.1|cs1A 68 3.5e-09
61 pp_1195|estExt_fggenesh1_pg.C_2130003|cesA 200 8.7e-75
64 os_53309|LOC_Os09g39920.1|cs1A 79 3.9e-06
66 os_53309|LOC_Os09g39920.1|cs1A 60 2.2e-09
69 sb_5917|estExt_Genewise1.C_chr_10425|cesA 145 9e-34
72 sm_4713|e_gw1.0.15.1|cesA 261 1.1e-115
73 os_42916|LOC_Os07g36630.1|cs1F 33 1.4e-07
74 pp_18179|e_gw1.373.38.1|cs1C 92 1.4e-23
76 pp_5273|estExt_gwp_gw1.C_1890065|cesA 140 1.3e-48
78 pt_53607|estExt_fggenesh1_pg_v1.C_LG_IX0974|cesA 150 4.7e-65
final-contig-data
  
```

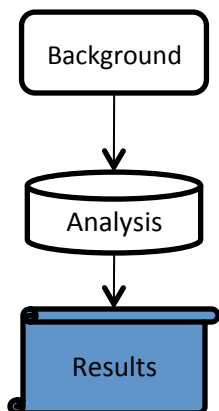
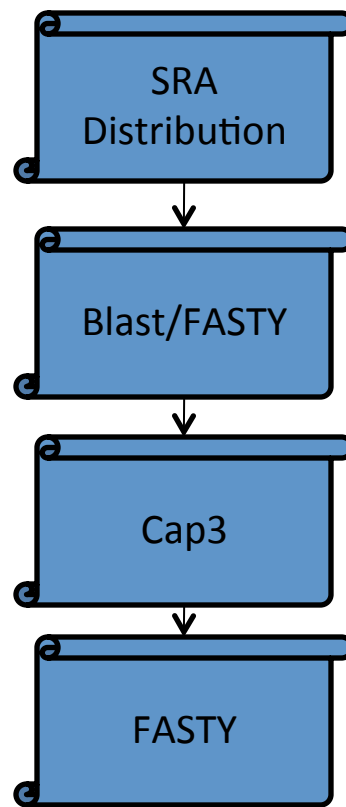




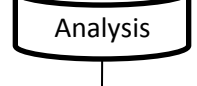
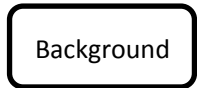
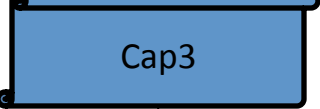
Analysis Stage	Sequences
SRA	614139
TBlastn/TFASTY	91126/94904



Analysis Stage	Sequences
SRA	614139
Blast/TFASTY	91126/94904
CAP3	4384

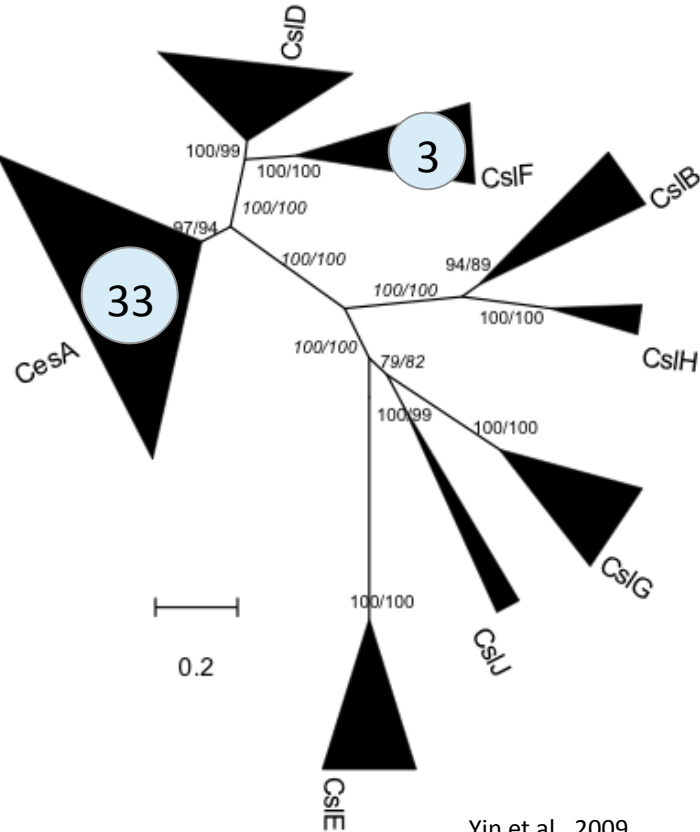


Analysis Stage	Sequences
SRA	614139
Blast/TFASTY	91126/94904
CAP3	4384
Final FASTY	74

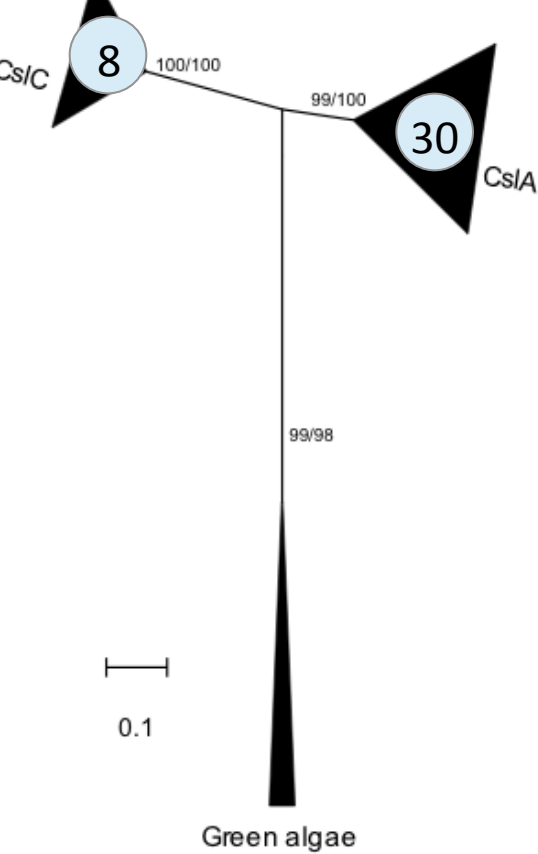


Family	Sequences
cesA	33
csIA	30
csIC	8
csIF	3
Total	74

A)

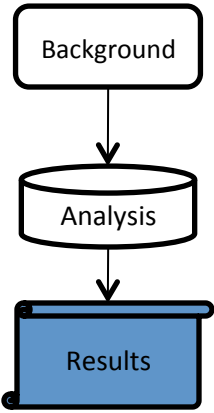
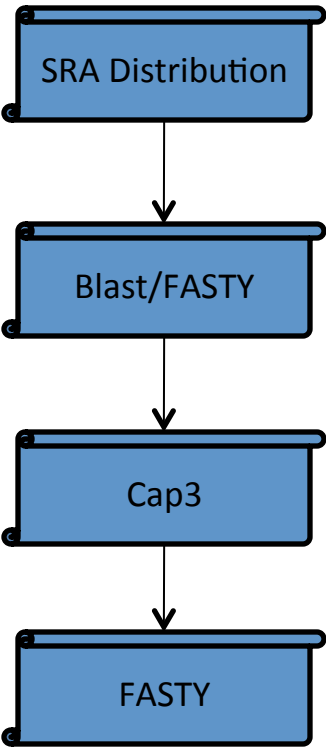


B)



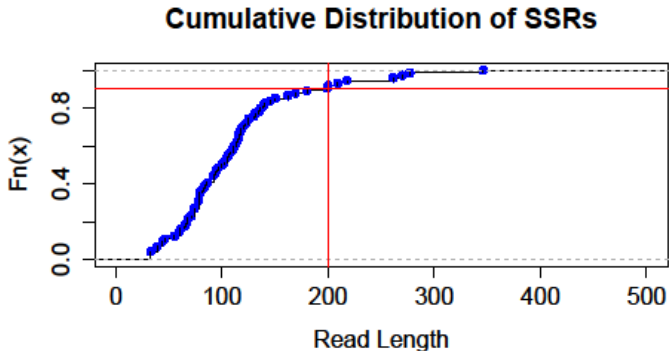
Yin et al., 2009

Green algae



	contig	x	y	family	length	eval
10	41pt_30735	eugene3.00160483		cesA	39	1.80E-12
67	375pt_55880	estExt_Genewise1_v1.C_LG_VI2190		cesA	44	8.50E-10
72	383sb_14834	estExt_Genewise1Plus.C_chr_106936		cesA	44	3.90E-06
25	82pt_30735	eugene3.00160483		cesA	55	1.80E-20
61	362sm_4713	e_gw1.0.15.1		cesA	60	1.90E-19
38	292os_5702	LOC_Os01g54620.1		cesA	66	1.90E-19
46	326pp_5273	estExt_gwp_gw1.C_1890065		cesA	68	1.70E-06
31	94pp_8706	estExt_fgenes1_pm.C_300004		cesA	83	3.00E-31
42	315pt_34236	eugene3.13980001		cesA	86	5.20E-32
5	31os_5702	LOC_Os01g54620.1		cesA	92	7.70E-36
66	370pt_30735	eugene3.00160483		cesA	94	4.40E-35
28	88pp_8706	estExt_fgenes1_pm.C_300004		cesA	95	2.70E-38
59	359pt_3118	fgenes1_pg.C_LG_V000269		cesA	97	4.60E-36
4	29pp_2722	estExt_Genewise1.C_80120		cesA	108	4.60E-29
62	363pp_8706	estExt_fgenes1_pm.C_300004		cesA	110	3.70E-47
39	305pp_4876	e_gw1.29.2.1		cesA	114	4.80E-42
34	100sm_4713	e_gw1.0.15.1		cesA	115	4.90E-40
11	45pp_5273	estExt_gwp_gw1.C_1890065		cesA	117	1.10E-26
32	95pp_8706	estExt_fgenes1_pm.C_300004		cesA	131	1.70E-58
6	33vv_19270	GSVIVP00025577001		cesA	135	2.60E-19
12	50vv_22004	GSVIVP00028800001		cesA	136	1.10E-49
7	34pp_8706	estExt_fgenes1_pm.C_300004		cesA	139	6.10E-63
22	76pp_5273	estExt_gwp_gw1.C_1890065		cesA	140	1.30E-48
18	69sb_5917	estExt_Genewise1.C_chr_10425		cesA	145	9.00E-34
23	78pt_53607	estExt_fgenes1_pg_v1.C_LG_IX0974		cesA	150	4.70E-65
60	360sm_4713	e_gw1.0.15.1		cesA	169	3.50E-55
26	83pp_5275	estExt_gwp_gw1.C_1890069		cesA	180	8.80E-39
69	377pp_4876	e_gw1.29.2.1		cesA	199	1.40E-88
15	61pp_1195	estExt_fgenes1_pg.C_2130003		cesA	200	8.70E-75
29	89vv_22004	GSVIVP00028800001		cesA	218	1.30E-95
19	72sm_4713	e_gw1.0.15.1		cesA	261	1.10E-115
9	37pp_4876	e_gw1.29.2.1		cesA	270	3.80E-63
1	18sm_18364	estExt_Genewise1Plus.C_490155		cesA	346	8.00E-105
54	344os_53309	LOC_Os09g39920.1		csIA	46	1.50E-07
17	66os_53309	LOC_Os09g39920.1		csIA	60	2.20E-09
33	97os_53309	LOC_Os09g39920.1		csIA	61	6.80E-09
14	60os_53309	LOC_Os09g39920.1		csIA	68	3.50E-09
24	81pt_41925	eugene3.00660037		csIA	71	2.40E-10
27	85os_53309	LOC_Os09g39920.1		csIA	74	2.40E-11
35	138os_53309	LOC_Os09g39920.1		csIA	74	1.30E-07
47	327os_53309	LOC_Os09g39920.1		csIA	74	2.60E-09
13	57os_53309	LOC_Os09g39920.1		csIA	78	1.60E-10
50	334os_53309	LOC_Os09g39920.1		csIA	78	9.80E-06
8	35os_53309	LOC_Os09g39920.1		csIA	79	1.90E-08
16	64os_53309	LOC_Os09g39920.1		csIA	79	3.90E-06
55	346os_53309	LOC_Os09g39920.1		csIA	79	6.50E-09
43	317vv_17281	GSVIVP00023244001		csIA	81	3.90E-12
44	318os_53309	LOC_Os09g39920.1		csIA	84	3.80E-09
73	385os_53309	LOC_Os09g39920.1		csIA	100	1.00E-08
45	320os_53309	LOC_Os09g39920.1		csIA	105	8.60E-09
51	335os_53309	LOC_Os09g39920.1		csIA	105	6.40E-12

53	342os_53309	LOC_Os09g39920.1	csIA	106	3.30E-06
52	341os_53309	LOC_Os09g39920.1	csIA	111	6.10E-12
74	389os_53309	LOC_Os09g39920.1	csIA	113	4.00E-12
58	358os_53309	LOC_Os09g39920.1	csIA	116	2.30E-10
64	368os_53309	LOC_Os09g39920.1	csIA	116	7.80E-13
37	288os_53309	LOC_Os09g39920.1	csIA	118	5.80E-07
68	376os_53309	LOC_Os09g39920.1	csIA	120	9.30E-09
63	367os_53309	LOC_Os09g39920.1	csIA	122	4.00E-10
30	90os_53309	LOC_Os09g39920.1	csIA	125	9.70E-11
56	348os_53309	LOC_Os09g39920.1	csIA	125	7.50E-06
41	310os_53309	LOC_Os09g39920.1	csIA	130	8.90E-14
40	308os_53309	LOC_Os09g39920.1	csIA	162	1.30E-18
65	369pp_1877	estExt_Genewise1.C_2240022	csIC	33	8.80E-10
57	354vv_15251	GSVIVP00020790001	csIC	65	3.60E-16
48	328pp_29689	estExt_gwp_gw1.C_190149	csIC	78	1.90E-13
3	25sb_5965	estExt_Genewise1.C_chr_11372	csIC	92	6.00E-38
21	74pp_18179	e_gw1.373.38.1	csIC	92	1.40E-23
71	380pp_29689	estExt_gwp_gw1.C_190149	csIC	102	2.30E-29
70	379sb_5965	estExt_Genewise1.C_chr_11372	csIC	209	9.50E-88
36	201vv_10325	GSVIVP00014999001	csIC	277	1.30E-15
2	23os_42916	LOC_Os07g36630.1	csIF	33	1.90E-07
20	73os_42916	LOC_Os07g36630.1	csIF	33	1.40E-07
49	329os_42916	LOC_Os07g36630.1	csIF	39	6.50E-06



References

Fangel, J. U., Ulvskov, P., Knox, J. P., Mikkelsen, M. D., Harholt, J., Popper, Z. A., & Willats, W. G. (2012). Cell wall evolution and diversity. *Frontiers in Plant Science*, 3, 152. doi:10.3389/fpls.2012.00152

Huang, X., & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Research*, 9(9), 868-877. doi:10.1101/gr.9.9.868

Spirogyra Pratensis Transeau :: *Algaebase*. (n.d.). [Web page]. Retrieved from www.algaebase.org:

http://www.algaebase.org/search/species/detail/?species_id=59533 Timme, R. E., & Delwiche, C. F. (2010). Uncovering the evolutionary origin of plant molecular processes: Comparison of coleochaete (coleochaetales) and spirogyra (zygnematales) transcriptomes. *BMC Plant Biology*, 10, 96. doi:10.1186/1471-2229-10-96

Timme, R. E., Bachvaroff, T. R., & Delwiche, C. F. (2012). Broad phylogenomic sampling and the sister lineage of land plants. *PloS One*, 7(1), e29696. doi:10.1371/journal.pone.0029696

Wodniok, S., Brinkmann, H., Glöckner, G., Heidel, A. J., Philippe, H., Melkonian, M., & Becker, B.

(2011). Origin of land plants: Do conjugating green algae hold the key? *BMC Evolutionary Biology*, 11, 104. doi:10.1186/1471-2148-11-104

Yin, Y., Huang, J., & Xu, Y. (2009). The cellulose synthase superfamily in fully sequenced plants and algae. *BMC Plant Biology*, 9, 99. doi:10.1186/1471-2229-9-99

Background

Analysis

Results

Contributions...

