

Supplementary Table I. – PCR primers used in Sato 2008 to produce the data set analysed here.

Marker	Name	Reference
<i>SSU</i>	A ^a	Medlin <i>et al.</i> (1988) ^c
	528F	Elwood <i>et al.</i> (1985)
	1055F	Elwood <i>et al.</i> (1985)
	536R	Elwood <i>et al.</i> (1985)
	1055R	Elwood <i>et al.</i> (1985)
	B ^b	Medlin <i>et al.</i> (1988) ^c
<i>LSU</i>	D1RF ^a	Scholin <i>et al.</i> (1994)
	D2CR ^b	Scholin <i>et al.</i> (1994)
<i>rbcL</i>	DPrbcL1 ^a	Daugbjerg & Andersen (1997)
	AraphidF ^a	This study ^d
	16F	Jones <i>et al.</i> (2005)
	14R	Jones <i>et al.</i> (2005)
	DPrbcL7 ^b	Daugbjerg & Andersen (1997)
<i>psbA</i>	psbA-F ^a	Yoon <i>et al.</i> (2002)
	psbA500F	Yoon <i>et al.</i> (2002)
	psbA-R2 ^b	Yoon <i>et al.</i> (2002)
	psbA600R	Yoon <i>et al.</i> (2002)

^a Forward PCR amplification primer.

^b Reverse PCR amplification primer.

^c Without polylinkers.

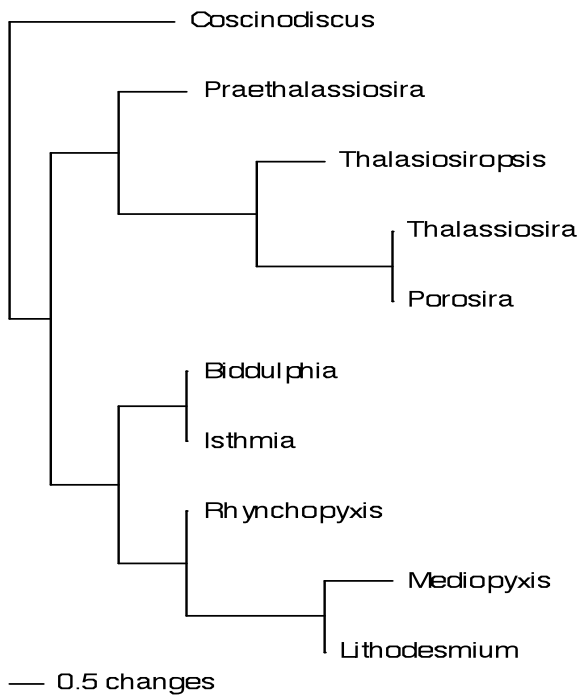
^d 5'-GTCTCAATCTGTATCAGAAC-3'.

Supplementary File I. – Cladistic analysis of morphological features of selected genera of the diatoms to test the position of *Praethalassiosira* and *Thalassiosiroopsis* among the centric diatoms.

CHARLABELS

- [1] position of cribrum in loculate areolum: 0 not loculate 1 external 2 internal,
- [2] position of tube process: 1 marginal 2 central 3 both
- [3] tube open or closed: 0 none 1 present open 2 present closed
3 some_open_some_closed,
- [4] satellite_pores_associate_with tube 1 absent 2 present,
- [5] VALVE_shape: 1 radial 2 'non-radial',
- [6] marginal_ridge_or_flange: 1 no 2 yes,
- [7] type_of_labiate: 0 none 1 macro_and_micro 2 macro 3 bilabiate_4 micro micro
- [8] type of areolation 1 radial 2 eccentric

Coscinodiscus	11111111
Praethalassiosira	22211102
Thalassiosiroopsis	22121122
Thalassiosira	23321122
Porosira	23321122
Mediopyxis	02312231
Lithodesmium	02112231
Rhynchopyxis	22112201
Biddulphia	22112141
Isthmia	22112141

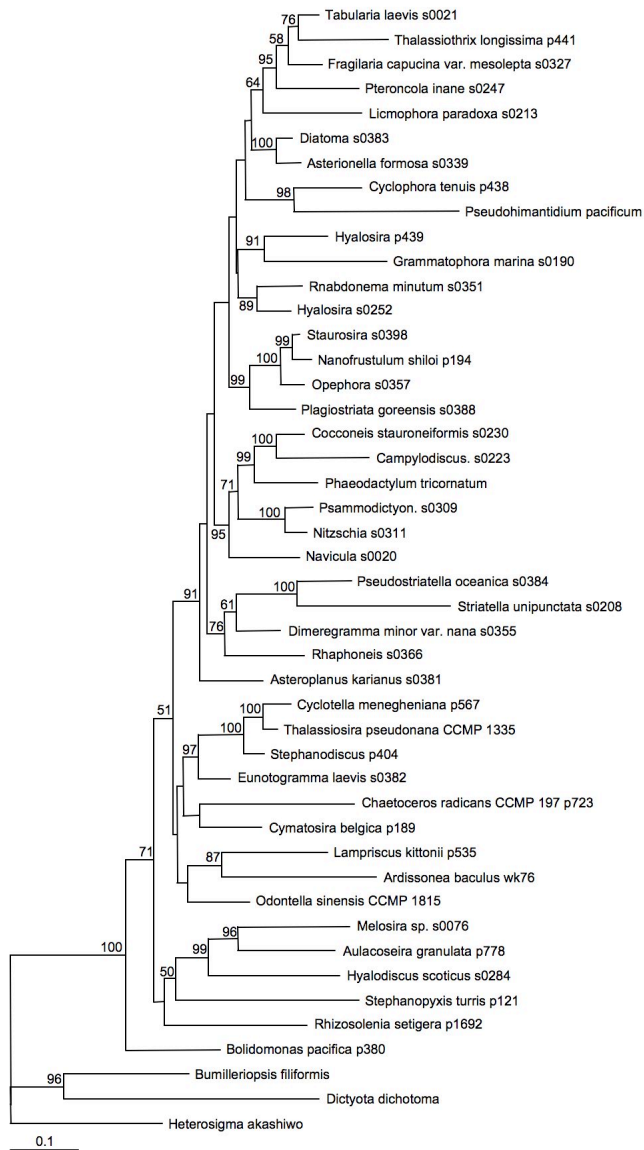


Supplementary File II. – Positions deleted from the alignment of the 18S SSU.

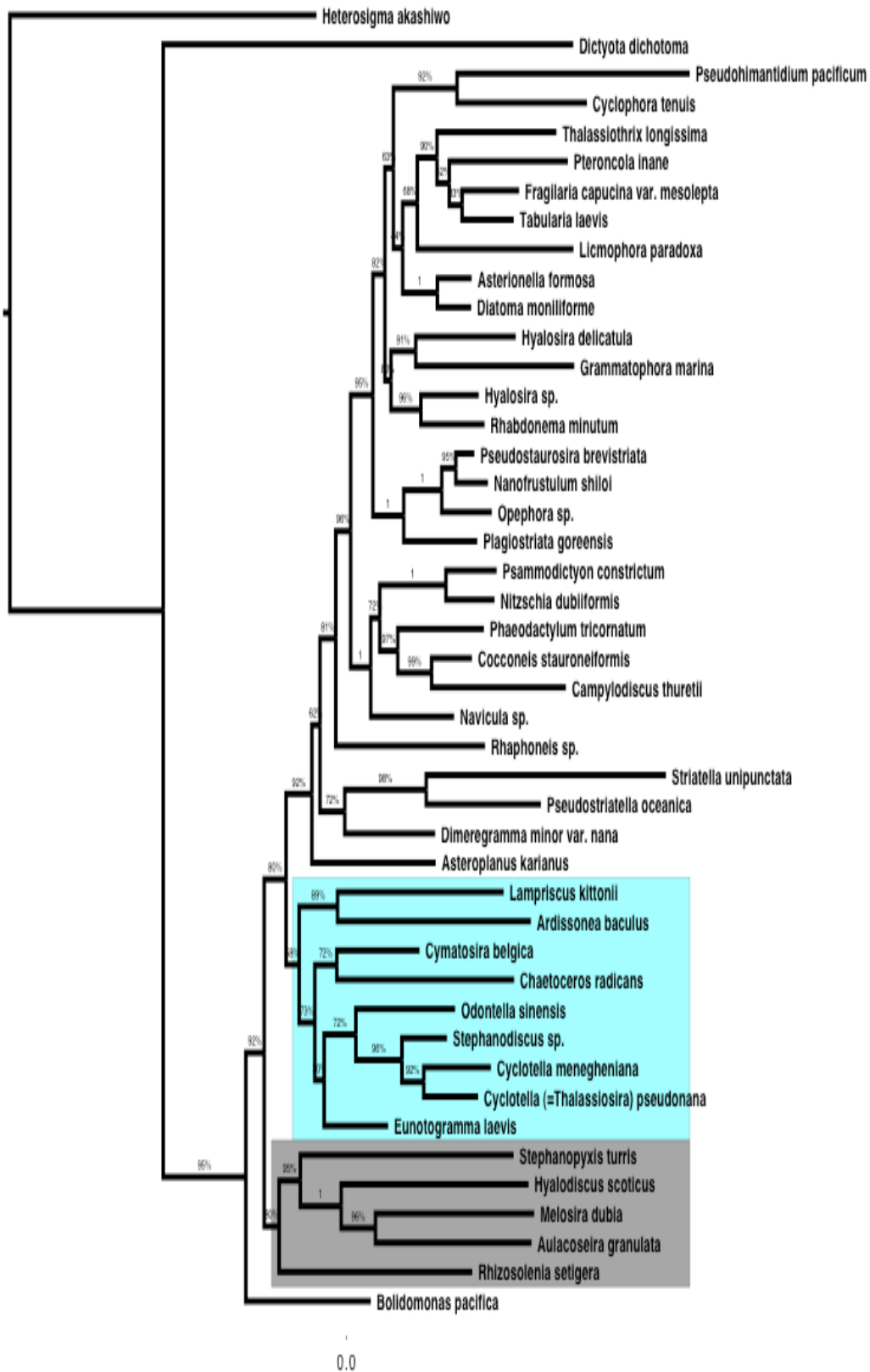
1-3 9-12 21 32 35 41 53 63 66 74 83-92 99 102 104 112 147-151 198 206-226 245-248 257 258 264-267 273-276 292 296
301 302 313 318-323 331 349 354 517 540-544 572 595 602 653 663 675 685 691 705-770 785 787-789 793 799 801 805
819 823-830 834 839 865 868 873 887 888 930 944 945 952 959 965 989 1042 1099 1160 1162 1165-1172 1175 1202 1203
1209 1225 1231 1238 1291 1360 1362 1365 1367 1370 1381 1384 1387 1413 1433 1452 1455 1473 1480-1485 1487 1494-
1498 1506-1515 1535 1537 1539-1543 1550 1632-1639 1657 1669 1670 1740 1745 1839 1847-1852 1857 1859 1866 1880

Supplementary Fig. I. – Bootstrap Tree for the ML analysis of the data set with multiple distant heterokont outgroups from Sato 2008. Third position of amino acids coded RY.

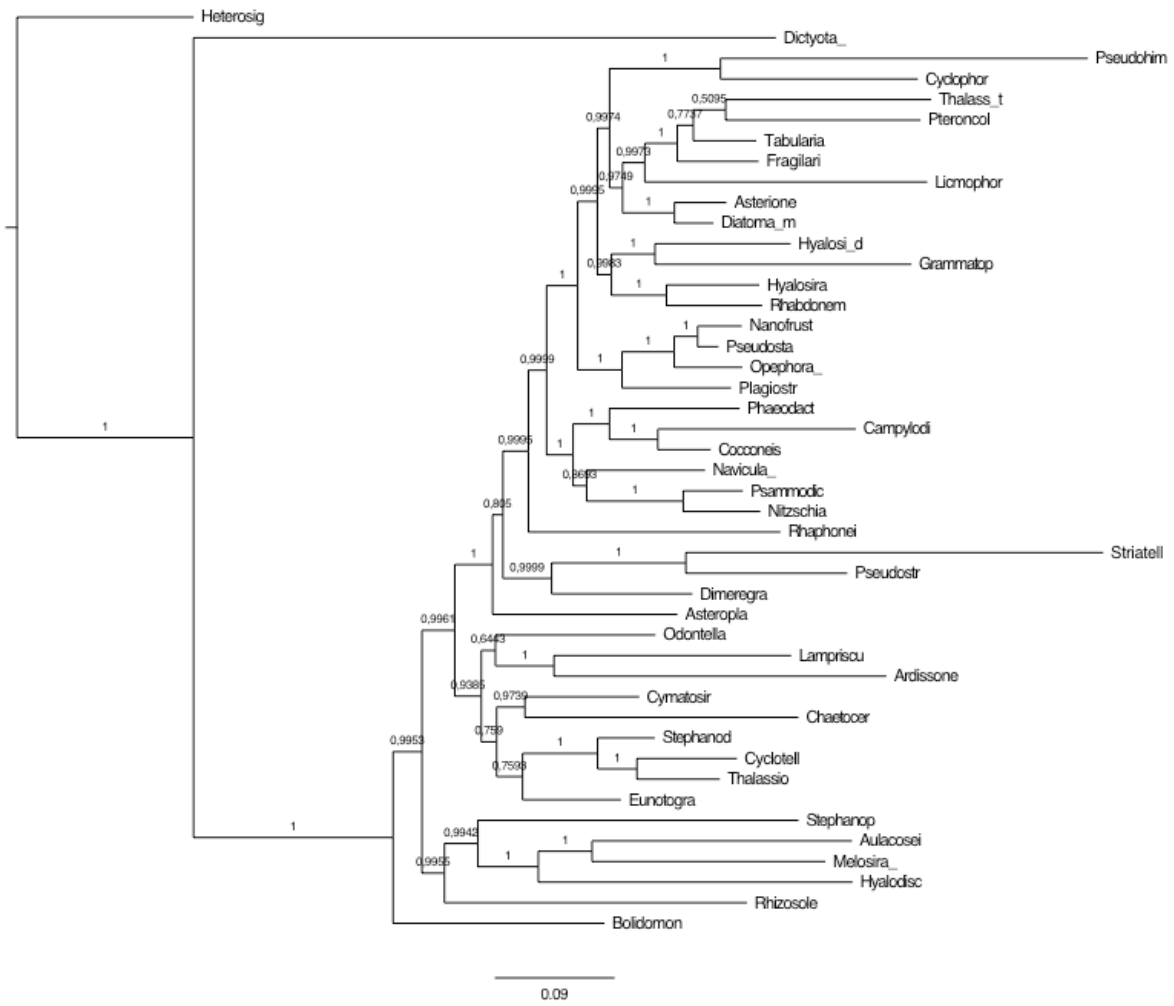
18S, 28S(RY), rbcL(3rd RY), psbA_ML x1000, bootstrap x 1000 24.12.2007



Supplementary Fig. 2. – Bootstrap Tree for the ML analysis of the data set with multiple distant heterokont outgroups from Sato 2008 without RY coding of the amino acid third position.



Supplementary Fig. 3. – Bootstrap Tree for the BI analysis of the data set with multiple distant heterokont outgroups from Sato 2008 without RY coding of the amino acids.



Supplementary Fig. 4. – Bootstrap Tree for the ML analysis of a data set with multiple bolidomonads outgroups and 157 diatoms using 3 genes. Data set from Li *et al.* 2015, except in this analysis the ARB alignment for the SSU and the LSU was used and this included the V4 region, which made the BT support higher than in Li *et al.* 2015.

