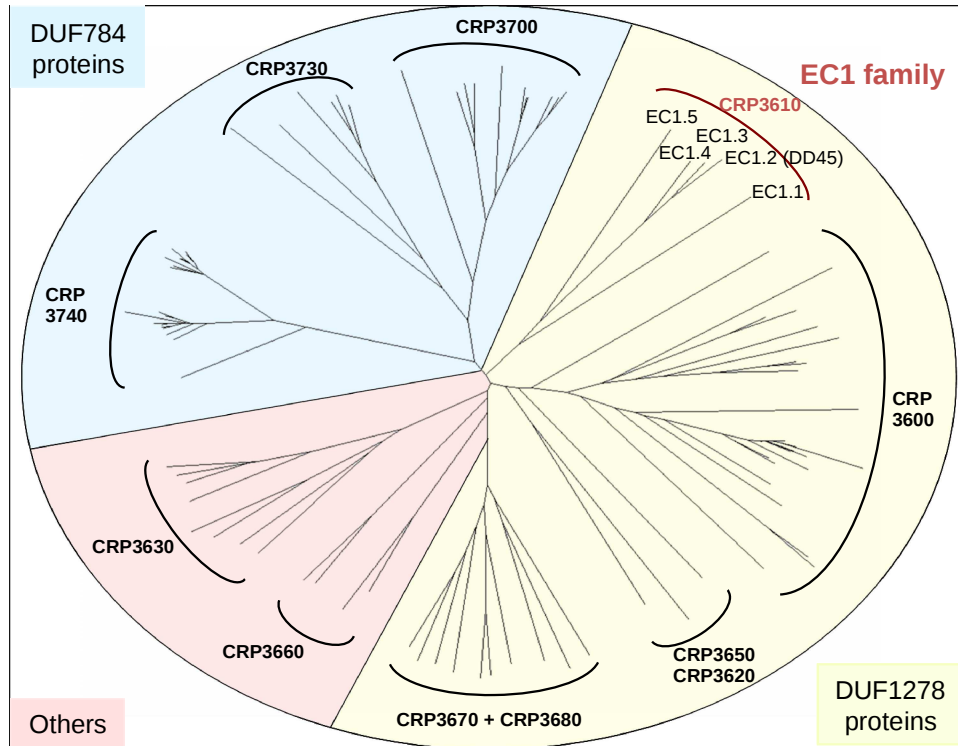


A



B

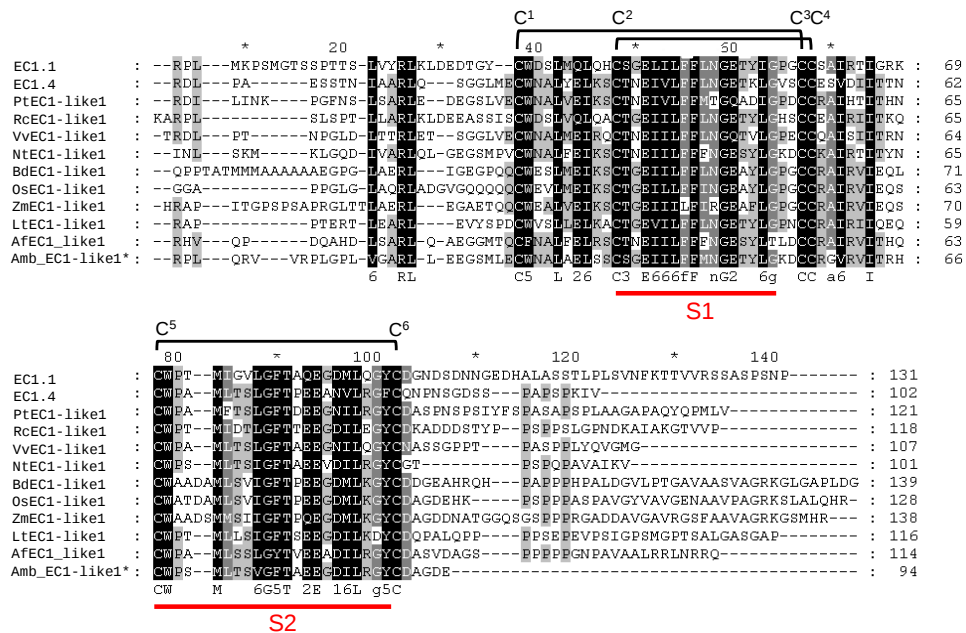


Figure S6.2 | The five *Arabidopsis* EC1 proteins form a distinct sub-clade within the large group of ECA1 gametogenesis-related cysteine-rich proteins (CRPs) and are char-

acterized by a conserved cysteine-spacing signature and by two conserved signature motifs.

(A) Predicted precursor protein sequences of 118 *Arabidopsis* genes, annotated as ECA1 gametogenesis-related CRPs (according to Silverstein et al. (2007) but excluding putative pseudogenes), were aligned using the program ClustalW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). The alignment was graphically rendered as an unrooted tree with Phylodraw, version 0.8 (<http://pearl.cs.pusan.ac.kr/phylodraw/>). The EC1 protein family of *Arabidopsis* belongs to the large subgroup of DUF1278-containing CRPs (yellow background), and has been classified as CRP 3619. One member of the gene family (*EC1.2*) was previously identified as *DD45* in a differential microarray screen (Steffen et al., 2007). Other ECA1 gametogenesis-related CRPs belong to the DUF784 proteins (blue background), or to the groups CRP3630, and CRP3660 (red background). For improved overview of the unrooted tree, the assigned AGI identifiers of the ECA1 gametogenesis-related CRPs are not shown. **(B)** The alignment of mature EC1.1 and EC1.4 with some of the EC1-like proteins identified in eudicots (Nt, *Nicotiana tabacum*; Pt, *Populus trichocarpa*; Rc, *Ricinus communis*; Vv, *Vitis vinifera*), monocots (Bd, *Brachypodium distachion*; Os, *Oryza sativa*; Zm, *Zea mays*), and basal angiosperms (Amb, *Amborella trichopoda*; Af, *Aristolochia fimbriata*; Lt, *Liriodendron tulipifera*) reveals two conserved EC1 signature motifs (S1, S2). Predicted N-terminal signal peptides for secretion were removed from the pre-proteins before alignment. Six cysteine residues (C¹ to C⁶) share a highly conserved spacing signature and are predicted to form three disulfide bonds (black brackets). Letters in black blocks label identical amino acid residues or conserved substitutions, amino acid residues with more than 80% conservation are highlighted in dark grey, and those with more than 60% conservation are shown as light grey boxes. Consensus sequences are given below the alignment. Signature motif 1 is located in a cys-loop, predicted to be formed by the intramolecular disulfide-bond between C² and C⁴, while signature motif 2 is part of the cys-loop generated by C⁵ and C⁶.

*note that EC1-like1 from *Amborella trichopoda* is derived from a draft sequence (Amborella Genome Project, www.amborella.org [12.01.2012]) and may not represent the complete protein.