J Plant Growth Regul (2005) 24:102–115
DOI: 10.1007/s00344-005-0033-y

J P G R
Journal of Plant Growth Regulation

EST Sequencing from Embryogenic *Cyclamen persicum* Cell Cultures Identifies a High Proportion of Transcripts Homologous to Plant Genes Involved in Somatic Embryogenesis

Stefan A. Rensing,[1] Daniel Lang,[1] Eik Schumann,[1] Ralf Reski,[1] and Annette Hohe[2]*

[1]*Plant Biotechnology, Faculty of Biology, University of Freiburg, Schaenzlestr, 1, D-79104 Freiburg, Germany;* [2]*Institute of Vegetable and Ornamental Crops, Kuehnhauser Str. 101, D-99189 Kuehnhausen, Germany*

ABSTRACT

To learn more about the process of somatic embryogenesis in the economically important ornamental plant *Cyclamen persicum*, we initiated an expressed sequence tag (EST) project. A normalized cDNA library was constructed from embryogenic cell material in different developmental stages, and clones were subsequently sequenced from the 5′ end. A total of 2083 filtered EST with an average length of 499 bases were analyzed in this study and submitted to the international sequence databases. By computational analyses, the *Cyclamen* transcripts were annotated and checked against plant genes previously described to be involved in somatic embryogenesis. Approximately one third of those genes were covered by the *Cyclamen* EST analyzed in this study. A high proportion of homologs to genes involved in somatic embryogenesis in the model system *Daucus carota* (carrot) were found in the *Cyclamen* EST collection. Of special interest are transcripts encoding gibberellin oxidases and somatic embryogenesis receptor-like kinases (SERK), both of which were confirmed to be important for development of embryos from somatic carrot cells. In addition, the set of candidate genes was expanded by using gene ontology (GO) annotations as well as by comparison with EST that were shown to be upregulated during *Glycine max* (soybean) somatic embryogenesis in a microarray approach. Our computational biology approach disclosed a set of around 90 candidate genes that now can be tested in the wet lab for their influence on somatic embryogenesis in *Cyclamen*. The annotated *Cyclamen* transcripts are available via www.Cyclamen-est.de.

Key words: *Cyclamen persicum*; Somatic embryogenesis; EST clustering

Received: 24 March 2005; accepted: 17 May 2005; Online publication: 30 September 2005

**Corresponding author; e-mail:* hohe@erfurt.igzev.de

Introduction

Cyclamen persicum is one of the most economically important ornamental potted plants in Europe. It is estimated that about 200 Million flowering plants are produced worldwide, from which at least 150 Million are grown in Europe (Schwenkel 2001). However, *Cyclamen* are one of the few ornamental potted plants with international economic importance that are exclusively propagated via seeds, because conventional vegetative propagation (for example, via cuttings) is not possible because of the plant habitus and regeneration characteristics.

Because a vegetative propagation system is highly desirable for multiplication of hybrid parents as well as selected elite genotypes, various *in vitro* propagation systems via organogenesis or somatic embryogenesis have been published (Winkelmann and others 2000). One of the most efficient systems is the protocol for somatic embryogenesis starting from unfertilized ovules (Schwenkel and Winkelmann 1998). This system displays many parallels to the well-known system of carrot somatic embryogenesis. Embryogenic callus develops from the explants on medium containing 2,4-dichlorophenoxyacetic acid (2,4-D) and 6-(gamma, gamma-dimethylallylamino) purine (2iP) as growth regulators. On the same medium, callus can be propagated by regular subculturing. Suspension cultures have been developed in liquid medium of the same composition (Winkelmann and others 1998). Moreover, automated propagation of the suspension cultures in stirred tank bioreactors has been established (Hohe and others 1999a, 1999b). As in carrot cell cultures, embryogenic *Cyclamen* suspensions contain so-called PEMs (proembryogenic masses), that is, aggregates of small cytoplasm-rich cells. By transfer of callus or suspension cultures to growth regulator–free medium differentiation of somatic embryos is initiated. This can be realized on solid media as well as in liquid medium (Hohe and others 2001). However, a major drawback of the system is a high number of developmentally aberrant embryos, insufficient maturation, and low desiccation tolerance (Seyring and Hohe 2005; Winkelmann and others 2004).

As of January 2005, a total of eight *Cyclamen persicum* sequences were available in the international sequence databases, the genus *Cyclamen* was represented by a mere 55 sequences. However, transformation protocols as a basis for molecular biology do exist (Aida and others 1999; Boase and others 2002). Because empirical studies on *in vitro* treatments promoting normal development and maturation were not sufficiently successful, we decided to establish an EST collection from cell material in different stages of embryogenesis as a basis for molecular work. This collection will enable us to study gene expression during embryo development and maturation as well as to identify candidate genes, whose relevance in somatic embryo- genesis can be studied by overexpression or knock-down.

Materials and Methods

Tissue Culture

The embryogenic cell line 03 was established from unfertilized ovules of a single plant from the cultivar 'Sierra Purple Flame' (Goldsmith Seeds, The Netherlands) in May 2003 (Schwenkel and Winkelmann 1998). The cell line was propagated by transfer to fresh medium every 4 weeks. Suspension cultures were established as described elsewhere (Winkelmann and others 1998). Realization of embryo development was induced by transfer of cells to plant growth regulator–free medium (Schwenkel and Winkelmann 1998). For isolation of RNA, the following cell material was collected: callus 14 days after subcultivation on plant growth regulator–containing medium, callus (including somatic embryos at different developmental stages) 1, 2, 3, 7, 14, and 21 days after transfer to growth regulator–free medium, and cells from a suspension culture 1 and 2 days after transfer to growth regulator–free medium. Both embryogenic callus as well as fully differentiated somatic embryos are shown in Figure 1.

Library Construction

The cDNA library was produced by Vertis Biotechnologie, Germany. Total RNA was isolated according to the method described by Chang and others 1993, poly(A)+ RNA was purified from 6.8 μg of total RNA by using Dynabeads Oligo(dT)25 (Dynal Biotech ASA, Oslo, Norway). During cDNA synthesis, oligonucleotide primers were attached to the 5′- and 3′-ends of the cDNA to allow polymerase chain reaction (PCR) amplification of the cDNA as well directional cloning of the cDNA into *Eco*RI/*Bam*HI-sites of pBluescript II SK(+). Normalization by denaturation/reassociation and subsequent separation of double- and single-stranded DNA using a hydroxylapatit column was carried out to equalize transcript abundance. The cDNA was size fractionated (>0.5 kbp) before cloning; transformation into *Escherichia coli* yielded 5.6 Million colony forming units. The number of non-recombinant clones was tested to be below 10%.

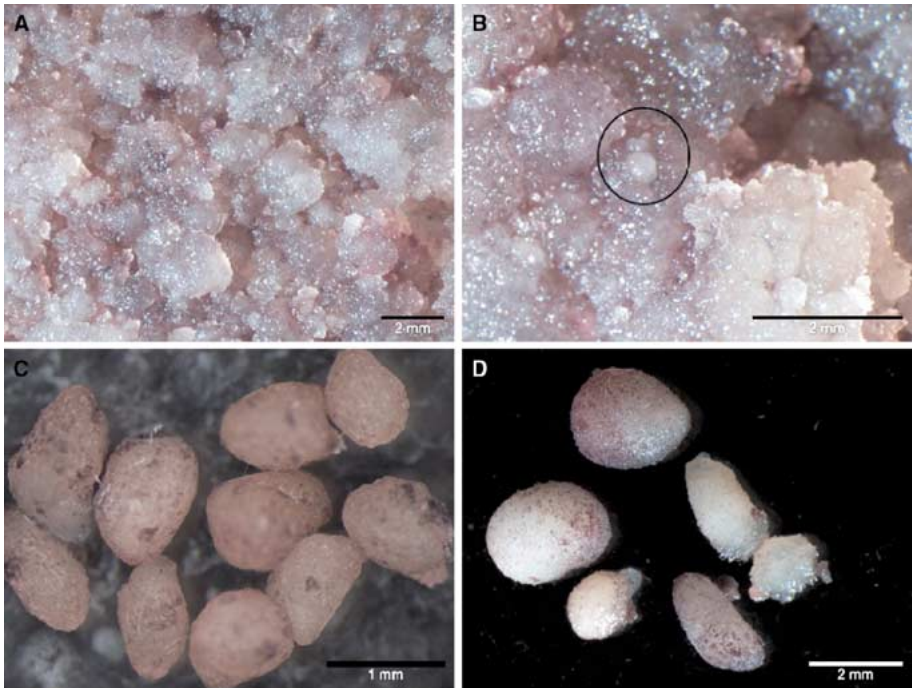

Figure 1. **A, B.** Embryogenic callus 21 days after subcultivation on medium containing growth regulators. Globular embryos differentiating on top of callus (not immersed in medium) are marked by a circle in b. **C.** Developing somatic embryos 14 days after transfer to growth regulator–free medium. **D.** Selected embryo stages 21 days after transfer to growth regulator–free medium.

Sequencing

Bacterial colonies were incubated in 250 µl LB medium containing 50 µg/µl ampicillin for 6 h at 37°C/120 rpm. DNA was prepared from 2 µl of this culture by using a TempliPhi amplification kit (Amersham Biosciences), according to the manufacturers protocol. Approximately 1.7 µg of circular DNA were used as a template for the sequencing reaction, which was carried out in 96-well plates with the CEQ DTCS Quickstart Sequencing kit (Beckman & Coulter). For each reaction, 5 pmol of an extended 34 mer T7 primer (5′- GTAATACGAC TCACTATAGG GCGAATTGGG TACC -3′) was used, that is, the 5′ end of the cDNAs was sequenced. Sequencing reactions were subjected to automated sequencing in a CEQ 8000 capillary laser fluorescence sequencer (Beckman & Coulter). The sequence raw data was then subjected to base calling and quality clipping using phred (Ewing and others 1998), with a trimming cutoff of 0.2. A total of 3230 clones were sequenced, yielding 3060 electropherograms and 2589 raw EST sequences (phred files).

EST Clustering and Annotation

Filtering, clustering, and assembly of EST data were carried out as described before (Lang and others 2005; Rensing and others 2002). During filtering, vector and linker regions were removed, and low complexity regions, untranslated regions (UTR), repeats, and poly-A/T were masked. After the final quality clipping and using a minimum length cutoff of 100 nucleotides, 2083 filtered input sequences with an average length of 499 bases remained, equalling 1.039 megabases. These *Cyclamen persicum* EST have been made available in the international sequence databases under the accession numbers AJ885882–AJ887964 and via www.cyclamen-est.de. The total number of sequences (assembled transcripts, that is, singlets and contigs) after clustering and assembly is 1980. Using EstScan 2.0 (Iseli and others 1999) with the *Arabidopsis thaliana* model, 1800 open reading frames (ORF) were predicted. The assembled transcripts were annotated using an annotation pipeline that makes use of a set of BLAST (Altschul and others 1997) and InterPro (Mulder and others 2003) searches and has been described before (Lang and others 2005). The pipeline also assigns Gene Ontology (GO) terms (Ashburner and others 2000) based on the GO associations (GOA) from the InterPro and UniProt databases (Camon and others 2003). The fully annotated *Cyclamen* transcripts were made available via www.cyclamen-est.de.

Detection of Candidate Transcripts Involved in Somatic Embryogenesis

Using NCBI (National Center for Biotechnology Information) Entrez (www.ncbi.nlm.nih.gov/entrez/), we created two data sets containing all publicly available *Viridiplantae* CDS (as of January 2005) associated with embryogenesis in general (623 sequences) and somatic embryogenesis in particular (237 sequences). BLAST hits (E-value cutoff 1E-3) of the *Cyclamen* transcripts against these data sets were added as similarity features to the annotated sequences.

In addition, we created a data set of 223 EST described to be upregulated during the development of somatic embryos in *Glycine max* (Thibaud-Nissen and others 2003) based on microarray experiments. Forty-nine of these soybean EST were found to be upregulated during the first 14 days of induction of somatic embryos on the adaxial side of cotyledons while the remaining 174 EST were upregulated 21 days and later. We used these sequences for BLAST searches against the predicted ORF of the assembled *Cyclamen* EST collection.

To further enhance the candidate population, we resolved the GOA for selected terms of all three ontologies. From the molecular function ontology (GO:0003674) the terms nucleotide binding (GO:0000166), nucleic acid binding (GO:0003676), receptor binding (GO:0005102), and receptor activity (GO:0004872) were used. From biological process (GO:0008150) the terms cell growth and/or maintenance (GO:0008151), signal transduction (GO:0007165), oxygen and reactive oxygen species metabolism (GO:0006800), regulation of metabolism (GO:0019222), response to external stimulus (GO:0009605), nucleosome assembly (GO:0006334), chromatin assembly or disassembly (GO:0006333), DNA methylation (GO:0006306), and histone deacetylation (GO:0016575)) were used. Finally, from cellular component (GO:0005575), the terms heterotrimeric G protein complex (GO:0005834), nucleus (GO:0005634), and chromosome (GO:0005694) were analyzed. For this selection of terms, we resolved the GOA for each of the terms by following the individual directed acyclic graphs down to the associated transcripts. The resulting set of transcripts was manually screened and filtered. The GO term associations of rice and *Arabidopsis* GOA projects (Rhee and others 2003; Ware and others 2002) were investigated via the AmiGO Browser (www.godatabase.org/cgi-bin/amigo/go.cgi as of February 2005). The three GOA used in this study show differences in terms of methodology, especially the annotation of gene products without a GO term association (for example, TAIR provides GOA for biological_process unknown (GO:0000004) – Gramene and our group do not). To compare the GOA directly, we did not allow GOA to the terms biological_process unknown (GO:0000004), cellular_component unknown (GO:0008372), and molecular_function unknown (GO:0005554).

Results and Discussion

Quality of the cDNA Library

We decided to normalize the cDNA to avoid redundancy by sequencing mainly highly abundant transcripts. From 2083 filtered sequences that went into the clustering and assembly, 1861 (89%) remained singlets. This high singlet rate is the result of efficient normalization. In addition, during clustering, no potential chimeras (cloning artifacts) were detected. We thus conclude that the library used is of high quality.

Annotation

Of 1980 putative transcripts, 1792 (90.1%) contained at least one predicted ORF. Nine hundred twenty-six transcripts (46.8%) could be assigned a high-quality *** annotation, for which the domain structure is considered by the comparison of the detected InterPro domain patterns on the query transcript and the BLAST-hits against the well-annotated UniProt (Apweiler and others 2004) database. For 57.6% of the sequences (1140 transcripts) a plant homolog could be detected. Interestingly, no BLAST hits in other taxonomic groups occurred. In total, 61.4% (1216) of the transcripts could be annotated.

Gene Ontology–based Classification of the Assembled Transcripts

By non-redundant mapping of the GO terms assigned in the annotation pipeline (Lang and others 2005), we classified the assembled transcripts in terms of molecular function, cellular component, and biological process. In Figure 2 we present the distribution of associated GO terms for the three basic Gene Ontology vocabularies (Ashburner and others 2000; Harris and others 2004). From the molecular function category (Figure 2a GO:0003674), we could assign 945 terms to 561 distinct transcripts, from biological process (Figure 2b GO:0008150) 608 terms to 478 distinct transcripts. From the cellular component category (Figure 2c GO:0005575), 301 terms could be assigned to 232 distinct transcripts. In total

1854 GO terms were assigned to 611 (30.8%) distinct transcripts.

The majority of terms assigned from the molecular function ontology belong to the catalytic activity category (GO:0003824; 50.1%) or the binding category (GO:0005488; 33.5%). This overall observation is comparable to the GO term distributions in *Arabidopsis* (GOA projects TAIR (Rhee and others 2003) and Tigr_Ath1) and rice (Gramene GOA [Ware and others 2002]) as they can be accessed via the AmiGO browser of the Gene Ontology Consortium (Ashburner and others 2000). This holds true also for the distribution of terms in the biological process ontology, where most of the terms (69.4%) are found in the metabolism category (GO:0008152). In the cellular component (GO:0005575) ontology, most terms were assigned from the categories protein complex (GO:0043234; 25.9%), nucleus (GO:0005634; 20.3%), and integral to membrane (GO:0016021; 12%). In comparison to the other plant GOA, a strong underrepresentation of associations from terms describing cytoplasmic localization (GO:0005737 cytoplasm; GO:0005829 cytosol) is obvious for our ''snapshot'' of the somatic embryogenesis transcriptome of *Cyclamen persicum*. In comparison to the TAIR GOA, where approximately 66% of the terms assigned for the cellular component belong to the category cytoplasm or one of its subcategories, in the *Cyclamen* GOA, this is the case for only 5.6%. Another category that is significantly under-represented in the *Cyclamen* GOA is organelle (GO:0043226). There are several categories that are over-represented as compared to the other plant GOA, like membrane (26% vs 17.8%), protein complex (25.9% vs 15.8%), or nucleus (20.3% vs 15.6%). Taken together, these findings indicate transcriptional differences between ''normal'' cells and those in a developing somatic embryo.

Transcripts Involved in Embryogenesis

Of the 1980 *Cyclamen* sequences, 116 found a BLAST homolog among the 860 plant proteins known to be involved in embryogenesis, equaling 13.5% of the search space. Although the hit rate among the 623 proteins associated with general embryogenesis was 5.9% (37 hits), there was a surprisingly high rate of hits against the plant proteins involved in somatic embryogenesis (79 hits, 33.3%). A third of the proteins described to play a role in somatic embryogenesis were covered by the *Cyclamen* genes presented in this work. We subsequently applied stringent filtering criteria to the BLAST hits to exclude potential false positive hits and reduced the redundancy by displaying only one EST per matched gene (Table 1).

Carrot Gene Homologs

The 38 non-redundant *Cyclamen* transcripts that matched a gene known to be involved in somatic embryogenesis are shown along with their annotation in Table 1. A total of 16 of the *Cyclamen* transcripts matched *Daucus carota* (carrot) genes, which were described to be important for somatic embryogenesis. Carrot is a well-studied model system in terms of somatic embryogenesis, and a lot of genes involved in somatic embryogenesis have been described from this organism (Chugh and Khurana 2002; Zimmerman 1993). Moreover, the carrot system is very similar to the *Cyclamen* somatic embryogenesis system. In both systems, development of PEMs (proembryogenic masses) in growth regulator–containing medium is described (Halperin 1966; Winkelmann and others 1998), and the realization of embryo development is initiated by transfer to growth regulator–free medium (Schwenkel and Winkelmann 1998; Zimmerman 1993).

Gibberellin

Of the 16 *Cyclamen* transcripts that matched *Daucus carota*, 8 were homologs of different isoforms of carrot gibberellin-oxidases (Table 1). Gibberellin is involved in the regulation of somatic embryogenesis in carrot (Tokuji and Kuriyama 2003). Specifically, the expression of the carrot gibberellin 3-oxidase transcript is induced, and biosynthesis of the plant hormone gibberellin is required during somatic embryogenesis in carrot (Mitsuhashi and others 2003). The development of aberrant phenotypes in carrot by application of uniconazole, an inhibitor of gibberellin biosynthesis, has been described (Tokuji and Kuriyama 2003), and this suggests that gibberellin may be required in later stages of somatic embryogenesis, especially in the differentiation of epidermal cells. This seems to be an interesting hint with regard to our *Cyclamen* somatic embryogenesis system, as a high proportion of the developing embryos display an aberrant phenotype (Seyring and Hohe 2005).

SERK

Another 7 *Cyclamen* transcripts matched genes annotated as SERK (somatic embryogenesis receptor–like kinase) from *Arabidopsis, Citrus, Daucus*, and *Oryza* (Table 1). This type of kinase has been shown to play a crucial role in somatic embryo-

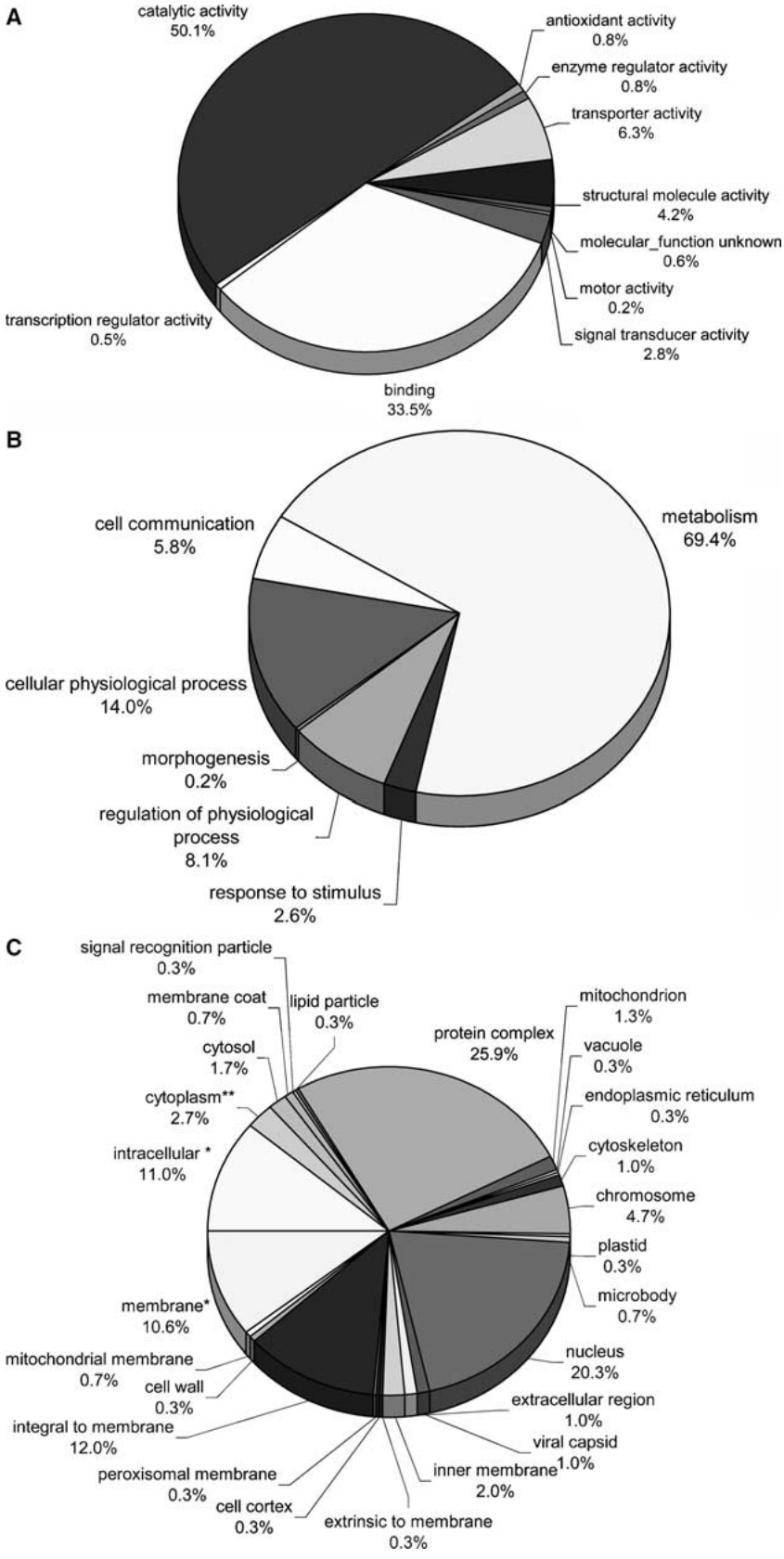

Figure 2. The distribution of associated GO terms in the *Cyclamen* GOA. Numbers given are the observed percentage frequencies of associations for the corresponding term. **A.** Molecular function category (GO:0003674), 945 terms were assigned to 561 distinct transcripts. **B.** Biological process category (GO:0008150), 608 terms were assigned to 478 distinct transcripts. **C.** Cellular component category (GO:0005575), 301 terms were assigned to 232 distinct transcripts. To increase the information content, we merged levels 3, 4, and 5 from the cellular component ontology.

Table 1. BLAST Hits against Proteins Involved in Somatic Embryogenesis

Gene product	Homolog involved in somatic embryogenesis	Acc.nr.	x hit	Cyclamen EST	Annotation
Calrèticulin	calreticulin [*Nicotiana plumbaginifolia*]	CAA95999	1	09T7.F03	***: (CRTC_NICPL) Calreticulin precursor.
Catalase	catalase [*Digitalis lanata*]	CAC04509	1	13T7.H06	***: (Q9SM65) Catalase (EC 1.11.1.6) (Fragment).
Cell wall protein	glycine-rich cell wall protein precursor - carrot	S56703	1	11T7.H04	***: (Q9M4H9) Putative proline-rich cell wall protein.
Chalcone synthase	chalcone synthase [*Digitalis lanata*]	CAA05512	5	12T7.A01	***: (Q8W3P5) Chalcone synthase (EC 2.3.1.74).
Chitinase	chitinase [*Picea glauca*]	AAA85364	2	12T7.E02	***: (Q9SDY6) Chitinase class I (EC 3.2.1.14).
Cysteine protease	cysteine protease [*Daucus carota*]	BAD29955	1	17T7.G09	***: (Q41690) Cysteinyl endopeptidase.
Cysteine protease	cysteine protease [*Daucus carota*]	BAD29956	1	31T7.E10	***: (Q6F6A7) Cysteine protease.
Cysteine protease	cysteine protease [*Daucus carota*]	BAD29958	2	33T7.F01	***: (Q5ZF64) Cysteine protease 1 (Fragment).
Cysteine protease	cysteine protease [*Daucus carota*]	BAD29960	1	26T7.C10	***: (Q84RM8) Papain-like cysteine proteinase isoform II.
Cytochrome P450	cytochrome P450 [*Pinus radiata*]	AAC05148	4	32T7.G09	***: (C718_MENPI) Cytochrome P450 71A8 (EC 1.14.-.-).
Gibberellin oxidase	gibberellin 20-oxidase1 [*Daucus carota*]	BAD30033	1	32T7.D12	***: (Q9SB32) SRG1-Mlike protein (At4g25310).
Gibberellin oxidase	gibberellin 20-oxidase2 [*Daucus carota*]	BAD30034	2	03T7.E12	***: (Q652U1) Putative 2-oxoglutarate-dependent oxygenase.
Gibberellin oxidase	gibberellin 3beta-hydroxylase2 [*Daucus carota*]	BAD30036	3	33T7.E07	***: (Q8HOBO) putative gibberellin 20-oxidase.
Gibberellin oxidase	gibberellin 2-oxidase2 [*Daucus carota*]	BAD30039	2	21T7.D10	***: (Q5QLCB) putative anthocyanidin synthase.
Glutathione S-transferase	Dcarg-1 [*Daucus carota*]	BAA78580	1	32T7.B11	***: (GTX1_TOBAC) Probable glutathione S-transferase (EC 2.5.1.18) (Auxin-induced protein PGNT1/PCNT110).
Glutathione S-transferase	glutathione S-transferase [*Cichorium intybus x Cichorium endivia*)	CAC24549	1	01T7.E12	***: (GTXA_TOBAC) Probable glutathione S-transferase parA (EC 2.5.1.18) (Auxin-regulated protein parA) (STR246C protein).
G-protein	GTP-binding protein [*Nicotiana plumbaginifolia*]	CAA69398	2	04T7.B01	***: (Q6EFV6) ARF-like small GTPase.
G-protein	small GTP-binding protein [*Nicotiana plumbaginifolia*]	CAA69701	4	31T7.G07	***: (Q8S2Z7) GTP-binding protein.
G-protein	GTP binding protein [*Cichorium intybus x Cichorium endivia*]	CAC24475	1	10T7.F10	***: (Q41668) Guanine nucleotide regulatory protein.
Homeobox	homeobox 1 [*Picea abies*]	AAG43405	1	30T7.G08	***: (O65281) Arabidopsis thaliana homeodomain protein AHDP (SP:P93041).
HSP18	heat shock protein 18 (clone pMsHsp18.1) - alfalfa (fragment)	S16248	1	14T7.G12	***: (Q9SSQ8) F6D8.22 protein (At1g52560).
kinase	receptor-like protein kinase [*Elaeis guineensis*]	AA026312	1	32T7.A03	***: (Q658G7) Putative transmembrane protein kinase.
kinase	receptor-like protein kinase [*Elaeis guineensis*]	AAO26313	1	04T7.B08	***: (Q9FXD7) F12A21.14.
Leucine-rich repeat protein	leucine-rich repeat family protein [*Arabidopsis thaliana*]	NP_189960	1	07T7.G10	***: (Q851U1) Putative leucine-rich repeat protein.
Lipase	lanatoside 15-O-acetylesterase [*Digitalis lanata*]	CAA09694	1	07T7.A11	***: (Q84K55) Putative family II extracellular lipase 1 (EXL1).
Peroxidase	peroxidase [*Asparagus officinalis*]	BAA94962	4	04T7.G04	***: (Q84UA9) Peroxidase 1.
PR protein	pathogenesis-related protein-like protein 2 [*Daucus carota*]	BAD04048	1	14T7.H03	***: (Q8L9L8) Pollen allergen-like protein.

Table 1. Continued

Gene product	Homolog involved in somatic embryogenesis	Acc.nr.	x hit	Cyclamen EST	Annotation
Secretory protein	embryo-abundant protein [*Picea glauca]*	AAB01559	1	22T7.A02	***: (Q9LV60) Putative 33 kDa secretory protein.
SERK	Similar to somatic embryogenesis receptor-like kinase [*Arabidopsis thaliana*]	AAD43169	2	13T7.C07	***: (O81833) Putative receptor protein kinase.
SERK	somatic embryogenesis receptor-like kinase, putative [*Arabidopsis thaliana*]	AAM63304	1	04T7.G06	***: (Q9LDG0) Putative receptor kinase (ESTs AU032341 (R3918)).
SERK	somatic embryogenesis receptor kinase 1 [*Citrus unshiu*]	BAD32780	1	03T7.G03	***: (Q6K4T4) Putative SERK2 protein.
SERK	putative benzothiadiazole-induced somatic embryogenesis receptor kinase 1 [*Oryza sativa* (*japonica cultivar-group*)]	BAD53863	1	12T7.B07	***: (Q5XWQ1) Serine/threonine protein kinase-like.
SERK	putative somatic embryogenesis protein kinase 1 [*Oryza sativa* (*japonica cultivar-group*)]	BAD68873	1	14T7.H01	***: (Q6DU55) S-locus-like receptor protein kinase (Fragment).
SERK	somatic embryogenesis receptor-like kinase-like protein [*Arabidopsis thaliana*]	CAB80060	1	26T7.C08	***: (Q69TY8) Receptor protein kinase-like.
Superoxide dismutase	manganese superoxide dismutase [*Digitalis lanata*]	CAC05259	1	25T7.A12	***: (O82583) Iron superoxide dismutase (EC 1.15.1.1).
WRKY	somatic embryogenesis related protein [*Dactylis glomerata*)	AAG42147	1	25T7.D02	***: (Q94IB3) WRKY DNA-binding protein.
WRKY	putative somatic embryogenesis related protein [*Oryza sativa japonica cultivar-group*)]	AAR87301	1	07T7.H10	***: (WR21_ARATH) Probable WRKY transcription factor 21 (WRKY DNA-binding protein 21).
Xyloglucan endotransglycosylase	putative xyloglucan endotransglycosylase [*Cucumis sativus*]	CAD88260	1	32T7.F01	***: (Q38696) Xyloglucan endotransglycosylase precursor.

BLAST hits of Cyclamen transcripts against plant proteins known to be involved in somatic embryogenesis. Hits were filtered according to these criteria: E-value < 1E-4, length of alignment > 50, identity > 25%. X hit = number of times each accession was hit. In cases where more than one hit occurred, only one of the Cyclamen expressed sequence tags (EST) that yielded the hits is shown, along with the result of the automated annotation.

genesis in carrot (Schmidt and others 1997), where the transcripts are expressed transiently in zygotic embryos as well as in competent somatic cells, demonstrating an embryogenesis-specific signal transduction chain mediated by SERK. High expression levels of SERK transcripts during embryogenesis and in somatic cells competent to form embryos were also found in other plants, for example, *Arabidopsis* and *Medicago* (Hecht and others 2001; Nolan and others 2003).

Candidate Selection Based on Gene Ontology

Using the GO annotation for determination of additional candidates involved in somatic embryogenesis, we analyzed the associated transcripts for selected terms of the molecular function, biological process, and cellular component ontology (see *Materials and Methods* for a list of terms). These queries led to another set of 40 promising candidate transcripts (excluding those already described in Table 1), which are presented in Table 2. Especially noteworthy are several transcription factors (MYB, HMG, zinc finger), cyclin-dependent kinase and ethylene-responsive factor, which might be part of specific signal transduction events culminating in the development of somatic embryos. Histone H3 and a histone deacetylase as well as a DNA methyltransferase hint at the activation of dormant genomic areas. Oxidative stress–related genes like peroxidases were found to be upregulated in the early phase of somatic embryo induction in microarray experiments in *Glycine max* (soybean) (Thibaud-Nissen and others 2003) and might be part of an oxidative burst during the first 14 days of 2,4-D treatment. Another hint supporting this assumption is the slight over-representation of terms from the oxygen and reactive oxygen species metabolism (GO:0006800) among the *Cyclamen* transcripts as compared to the genome-based GOA from *Arabidopsis* and rice. Although *Cyclamen* transcripts in the other metabolism subcategories are under-represented, the oxygen and reactive oxygen species metabolism category for the *Cyclamen* GOA is 2.5 times over-represented compared with the other plant GOA (TAIR, Tigr_Ath1, Gramene; data not shown).

Comparison with Soybean Micro array Data

By comparison with a set of 223 EST that were found to be upregulated during the induction and development of globular somatic embryos in soybean (Thibaud-Nissen and others 2003) using a micro array approach, we could detect an additional set of candidate transcripts. The query set was divided into 49 EST upregulated in the early phases of somatic embryo induction (set I; 0–14 days of auxin treatment) and 174 EST that were upregulated in the later phase of somatic embryo development (set II; 21–28 days of auxin treatment). A small percentage (6.1%) of the sequences from set I were covered by a homologous *Cyclamen* transcript whereas the percentage in set II was 13.8%. The lower coverage in the first set (induction phase) probably occurred because our RNA was prepared from later stages of somatic embryo development, which are comparable to set II (developmental phase) of the soybean data. Those 16 *Cyclamen* transcripts that were not already covered by the analyses shown above (Table 1 and Table 2) are presented in Table 3. Among the *Cyclamen* homologs of genes expressed in the early phase of soybean somatic embryogenesis (set I), phosphatases are of special interest, which might be involved in signal transduction cascades. Two peroxidases as indicators of a putative oxidative burst in response to the auxin treatment during the early phases of induction were also determined but are not displayed in Table 3 because they were already covered and discussed in the GO-based approach. In the set of transcripts expressed after 21 days and later (set II), there are auxin-induced gene products and histone H2A/H3, as well as a kinase and a phosphatase inhibitor. Furthermore, cysteine proteinases could be related to a possible accumulation of storage proteins, which was also described for the developing globular embryos after 14 days in soy-bean (Thibaud-Nissen and others 2003).

Candidate Genes to Influence *Cyclamen* Somatic Embryogenesis

The above-mentioned genes from the homology-based approach, the GOA, and the microarray data have been grouped into five classes and are presented in Figure 3. These transcripts are especially promising candidates for involvement in *Cyclamen* somatic embryogenesis. Many of the genes usually associated with embryogenesis and somatic cells becoming competent (Chugh and Khurana 2002; Thibaud-Nissen and others 2003) have thus been identified from *Cyclamen*.

Conclusions

By using a normalized library derived from tissue related to somatic embryogenesis, we were able to assign 4% of the EST to gene products that are

Table 2. Candidates Selected by Gene Ontology

Term	Term annotation	Accession	Annotation	n found
GO:0003677	DNA binding	13T7.C04	***: (H3_ARATH) Histone H3.	7
GO:0003677	DNA binding	16T7.G04	***: (Q8W1W6) Putative transcription factor MYB33 (Putative MYB family transcription factor) (MYB transcription factor).	7
GO:0003677	DNA binding	17T7.H04	***: (Q7XIZ8) Hislone H3 like protein.	7
GO:0003677	DNA binding	19T7.B03	contains: HMG-I and HMG-Y, DNA-binding (InterPro:IPR000637, GO:0003677, GO:0005634, GO:0006355, PRINTS:PR00929)	7
G0:0003677	DNA binding	19T7.C09	contains: HMG-I and HMG-Y, DNA-binding (InlerPro:IPR000637, GO:0003677, GO:0005634, GO:0006355, PRINTS:PR00929)	7
GO:0003677	DNA binding	29T7.F04	***: (Q6A332) Always early protein 3.	7
GO:0003677	DNA binding	32T7.E01	contains: Site-specific DNA-methyltransferase (cytosine-N4-specific) (InterPro:IPR001091, GO:0003677, GO:0006306, GO:0008170, PROSITE:PS00093)	7
GO:0003700	transcription factor activity	08T7.E01	***: (Q6S3EO) Homeodomain protein HB2 (Fragment).	2
GO:0003700	transcription factor activity	18T7.01	***: (QBQQS5) Ethylene-responsive factor-like protein 1.	2
GO:0003723	RNA binding	30T7.D11	***: (Q84LLB) Salt tolerance protein 4.	1
GO:0003746	translation elongation factor activity	12T7.H11	***: (Q8H2U3) Putative translation elongation factor eEF-1 beta chain.	1
GO:0004872	receptor activity	09T7.AO3	***: (Q9M7AB) LRR receptor-like protein kinase.	3
GO:0004872	receptor activity	26T7.H10	***: (O65462) Receptor like protein (Fragment).	3
GO:0004872	receptor activity	30T7.B03	***: (Q9FH56) Similarity to elicitor-inducible receptor-like protein (Hypothetical protein At5g66330).	3
G0:0004970	ionotropic glutamate receptor activity	01T7.B02	***: (Q69L05) Putative Avr9/Cf-9 rapidly elicited protein 141.	1
GO:0005184	neuropeptide hormone activity	08T7.B04	***: (Q62103) Proline-rich protein precursor.	1
GO:0005524	ATP binding	27T7.A06	***: (Q93X47) Cyclin dependent kinase C.	3
GO:0005524	ATP binding	18T7.E06	***: (O49409) Protein kinase-like protein.	3
GO:0005524	ATP binding	20T7.D06	***: (Q8VWW7) Putative receptor-like protein kinase RLPK1 (Fragment).	3
GO:0005525	GTP binding	16T7.F09	***: (Q6EFV6) ARF-like small GTPase.	3
GO:0005525	GTP binding	16T7.F11	***: (Q9C923) Putative GTP-binding protein; 106556-109264.	3
GO:0005525	GTP binding	17T7.G04	***: (O24110) Small GTP-binding protein.	3
GO:0005634	nucleus	13T7.G07	***: (Q9STM3) Putative zinc finger protein (Relative of early flowering 6).	1

Table 2. Continued

term	Term annotation	accession	annotation	n found
GO:0005834	heterotrimeric G-protein complex	13T7.E07	***: (Q6EE07) Notchless-like protein.	1
GO:0006306	DNA methylation	22T7.C09	contains: N-B Adenine-specific DNA methylase (InterPro:IPR002052, GO:0003677,GO:0006306, GO:0008170, PROSITE:PS00092)	1
G0:0006355	regulation of transcription, DNA-dependent	01T7.D02	***: (Q8LDB1) Putative transcription factor.	1
GO:0006955	immune response	19T7.A04	contains: Interleukin-4/interleukin-13(InterPro:IPR001325, GO:0005126, GO:0005576, GO:0006955, PROSITE:PS00838)	1
GO:0007001	chromosome organization and biogenesis (sensu Eukaryota)	31T7.D06	**: Homolog of (9550.m00130) histone H3 - maize|o_sativa|chr_11|OSJNBb0076K20|9550	1
GO:0007600	sensory perception	25T7.F10	***: (Q71BZ1) Type-B response regulator.	2
GO:0007600	sensory perception	18T7.G11	***: (Q9FT60) Histidine kinase-like protein.	2
G0:0009621	response to pathogenic fungi	32T7.D09	***: (O80630) Putative disease resistance response protein (At2g39430).	1
GO:0016575	histone deacetylation	20T7.CO1	***: (Q9FML2) Histone deacetylase.	1
GO:0006979	response to oxidative stress	21T7.G03	***: (Q9XFI6) Peroxidase.	7
GO:0006979	response to oxidative stress	02F.D03	***: (Q07446) Peroxidase precursor (EC 1.11.1.7).	7
GO:0006979	response to oxidative stress	04T7.G04	***: (Q84UA9) Peroxidase 1.	7
GO:0006979	response to oxidative stress	O9T7.E10	***: (PE11_ARATH) Peroxidase 11 precursor (EC 1.11.1.7) (Atperox P11) (ATP23a/ATP23b).	7
GO:0006979	response to oxidative stress	11T7.B02	***: (PE52_ARATH) Peroxidase 52 precursor (EC 1.11.1.7) (Atperox P52) (ATP49).	7
G0:0006979	response to oxidative stress	13T7.H06	***: (Q9SM65) Catalase (EC 1.11.1.6) (Fragment).	7
GO:0006979	response to oxidative stress	21T7.E12	***: (Q9XIV8) Peroxidase (EC 1.11.1.7).	7
GO:0006801	superoxide metabolism	25T7.A12	***: (O82583) Iron superoxide dismutase (EC 1.15.1.1).	1

Utilizing the GO annotation for the determination of additional candidate transcripts: For a selection of interesting GO terms (see Materials and Methods for a list), the complete set of associated transcripts was recovered. The resulting set of transcripts was manually screened for gene products with putative involvement in somatic embryogenesis. Only those candidate transcripts not already covered by the analysis shown in Table 1 were considered.

Table 3. Candidates Selected Using Soybean Microarray Data

Gene product	Annotation of soybean hit sequence	Soybean EST	x hit	Cyclamen EST	Annotation
Phosphatase	B′ regulatory subunit of PP2A	BE658565	2	29T7.B02	***: (Q9LU89) Protein phosphatase 2A regulatory subunit B′ (AT3g26020/MPE11_17).
Cysteine Proteinase	cysteine proteinase	BE824348	2	20T7.B05	***: (Q84RM8) Papain-like cysteine proteinase isoform II.
Auxin responsive protein	GH1/IAA9	BE657627	3	03T7.A01	***: (Q8RW16) Aux/IAA protein.
Chloroplast DNA-binding protein	Ring3-like bromodomain protein	AI748092	1	09T7.E03	***: (O04698) Chloroplast DNA-binding protein PD3.
ATP-binding protein	similarity to ATP-bd protein associated with cell differentiation	BE823658	1	11T7.B01	***: (Q8LCV1) ATP-binding protein-like protein (Hypothetical protein At5g66410).
Histine H2A	histone H2A.F/Z	BE823365	1	13T7.B04	***: (Q9C944) Putative histone H2A; 14481-15293 (Putative histone H2A) (At1g52740/F14G24_1).
Histine H3	histone H3	BE821098	2	13T7.C04	***: (H3_ARATH) Histone H3.
Kinase	uridylate kinase	BE823241	1	28T7.E01	***: (Q7XI40) UMP/CMP kinase a.
Argonaute	argonaute (zwille, pinhead) like protein	BE657421	1	31T7.E01	***: (Q9SDG8) ESTs AU068544 (C30430).
HMG protein	probable high mobility group protein HMG1	AI966779	1	32T7.D01	***: (Q9LM85) F2D10.18.
Phosphatase inhibitor	putative phosphatase 2A inhibitor	AW568120	1	33T7.F08	***: (Q9M9V0) F6A14. 10 protein.

BLAST hits of Cyclamen transcripts against a set of soybean EST, which are upregulated during somatic embryogenesis. Hits were filtered according to these criteria: E-value < 1E-4, length of alignment > 50, identity > 25%, X hit = number of times each accession was hit. In cases where more than one hit occurred, only one of the Cyclamen EST that yielded the hits is shown, along with the result of the automated annotation. The search space was divided into 49 EST upregulated in the early phases of somatic embryo induction (set I; 0–14 days of auxin treatment; above thick line) and 174 EST that were upregulated in the later phases of somatic embryo development (set II; 2–28 days of auxin treatment; below thick line). Only candidate transcripts not already covered by the analyses shown in Table 1 and Table 2 were considered.

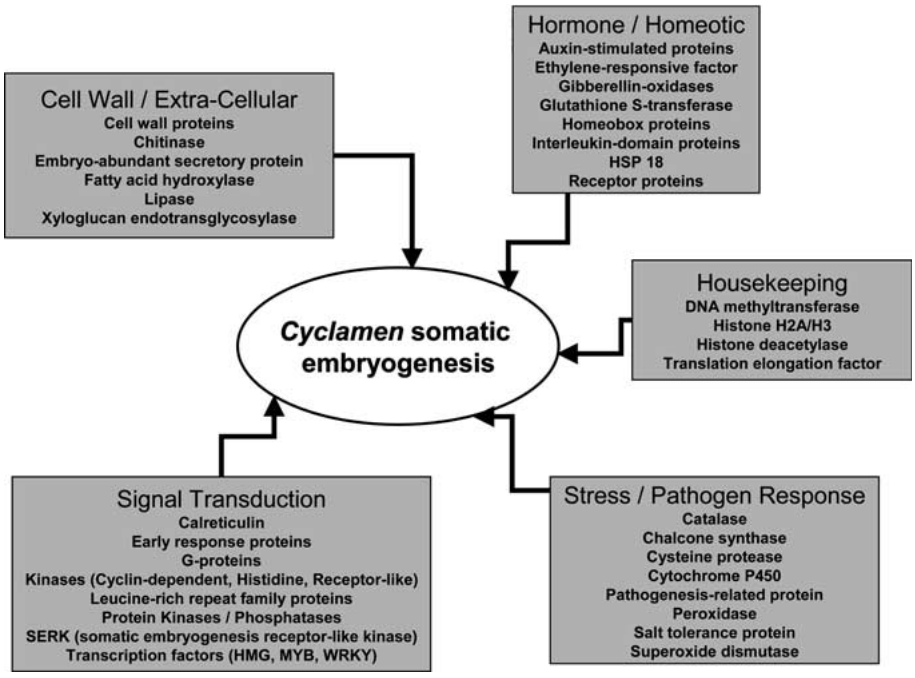

Figure 3. *Cyclamen persicum* genes putatively involved in somatic embryogenesis. ''Cell Wall/Extra-Cellular'' contains gene products that act at the cell wall or are secreted. ''Hormone/Homeotic'' indicates genes regulated by plant hormones or involved in hormone production/homeotic regulation. ''Signal Transduction'' indicates participation in signaling cascades. ''Stress/Pathogen Response'' represents genes that are expressed under stress conditions / as pathogen response and gene products involved in detoxification/protection.

putatively involved in the development of somatic embryos. *Cyclamen* homologs for approximately one third of all plant genes currently known to play a role in somatic embryogenesis were found.

The bulk of genes known to be key players in carrot somatic embryogenesis could be identified. The GA hormone pathway seems to be important during somatic embryogenesis of *Cyclamen persicum*. Several SERK transcripts were detected, pointing to a specific signal transduction pathway during genesis of embryos from somatic *Cyclamen* cells. Using gene ontology associations, the set of 38 candidate genes determined by the homology approach could be expanded by another 40, adding transcription factors, G proteins, proteins involved in reactive oxygen species, and others to the set. Another 11 *Cyclamen* candidate genes were discovered by comparison with data from soybean microarray experiments, including phosphatases, kinases, and histones.

Around 90 genes that may be involved in *Cyclamen* embryo development can now be checked for their influence on somatic embryogenesis. Transformation protocols for *Cyclamen* have recently been published (Aida and others 1999; Boase and others 2002), and Agrobacterium-mediated transformation of our *Cyclamen* embryogenic cell lines has already been positively tested (data not shown). This will allow for the study of the role of specific candidate genes by overexpression. Further analysis of specific transcript expression may be carried out using micro-array experiments, which have already been shown to be useful in soybean (Thibaud-Nissen and others 2003).

ACKNOWLEDGMENTS

The authors thank J. Eisinger for technical assistance with the annotation pipeline.

REFERENCES

Aida R, Hirose Y, Kishimoto S, Shibata M. 1999. *Agrobacterium tumefaciens* mediated transformation of *Cyclamen persicum* Mill. Plant Sci 148:1–7.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, others. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.

Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, others. 2004. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res 32 (Database issue):D115–D119.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, others. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25–29.

Boase M, Marshall G, Peters T, Bendall M. 2002. Long-term expression of the gusA reporter gene in transgenic *Cyclamen* produced from etiolated hypocotyls explants. Plant Cell Tiss Org Cult 70:27–39.

Camon E, Magrane M, Barrell D, Binns D, Fleischmann W, others. 2003. The Gene Ontology Annotation (GOA) project: implementation of GO in SWISS-PROT, TrEMBL, and InterPro. Genome Res 13:662–672.

Chang S, Puryear J, Cairney J. 1993. A simple and efficient method for isolating RNA from pine trees. Plant Mol Biol Rep 11:113–116.

Chugh A, Khurana P. 2002. Gene expression during somatic embryogenesis—recent advances. Current Sci 83:715–730.

Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8:175–185.

Halperin W. 1966. Alternative morphogenetic events in cell suspensions. Am J Bot 53:443–451.

Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, others. 2004. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 32 (Database issue):D258–D261.

Hecht V, Vielle-Calzada JP, Hartog MV, Schmidt ED, Boutilier K, others. 2001. The Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1 gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. Plant Physiol 127:803–816.

Hohe A, Winkelmann T, Schwenkel H. 1999a. CO_2 accumulation in bioreactor suspension cultures of *Cyclamen persicum* Mill. and its effect on cell growth and regeneration of somatic embryos. Plant Cell Rep 18:863–867.

Hohe A, Winkelmann T, Schwenkel H. 1999b. The effect of oxygen partial pressure in bioreactors on cell proliferation and subsequent differentiation of somatic embryos of *Cyclamen persicum*. Plant Cell Tiss Org Cult 59:39–45.

Hohe A, Winkelmann T, Schwenkel H. 2001. Development of somatic embryos of Cyclamen persicum Mill. in liquid culture. Gartenbauwissenschaft 66:219–224.

Iseli C, Jongeneel CV, Bucher P. 1999. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. In: Lengauer T, Schneider R, Bork P, Brutlag DL, Glasgow JI, others. International Conference on Intelligent Systems for Molecular Biology. Menlo Park, CA, American Association for Artificial Intelligence, p 138–148.

Lang D, Eisinger J, Reski R, Rensing SA. 2005. Representation and high-quality annotation of the *Physcomitrella patens* transcriptome demonstrates a high proportion of proteins involved in metabolism among mosses. Plant Biol 7:238–250.

Mitsuhashi W, Toyomasu T, Masui H, Katho T, Nakaminami K, others . 2003. Gibberellin is essentially required for carrot (*Daucus carota* L.) somatic embryogenesis: dynamic regulation of gibberellin 3-oxidase gene expressions. Biosci Biotechnol Biochem 67:2438–2447.

Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Barrell D, others. 2003. The InterPro Database, 2003 brings increased coverage and new features. Nucleic Acids Res 31:315–318.

Nolan KE, Irwanto RR, Rose RJ. 2003. Auxin up-regulates MtSERK1 expression in both *Medicago truncatula* root-forming and embryogenic cultures. Plant Physiol 133:218–230.

Rensing SA, Rombauts S, Van de Peer Y, Reski R. 2002. Moss transcriptome and beyond. Trends Plant Sci 7:535–538.

Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, others. 2003. The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic Acids Res 31:224–228.

Schmidt ED, Guzzo F, Toonen MA, de Vries SC. 1997. A leucine-rich repeat containing receptor-like kinase marks somatic plant cells competent to form embryos. Development 124:2049–2062.

Schwenkel H. (2001). Introduction: botany—economic importance—cultivars—micropropagation of C. persicum. In: Schwenkel H, editor . Reproduction of *Cyclamen persicum* Mill. through somatic embryogenesis using suspension culture systems. (EUR 19697). Publication of the European Commission.

Schwenkel H, Winkelmann T. 1998. Plant regeneration via somatic embryogenesis from ovules of *Cyclamen persicum* Mill. Plant Tiss Cult Biotechnol 4:28–34.

Seyring M, Hohe A. 2005. Induction of desiccation-tolerance in somatic embryos of *Cyclamen persicum* Mill. J Hort Sci Biotechnol 80:65–69.

Thibaud-Nissen F, Shealy RT, Khanna A, Vodkin LO. 2003. Clustering of microarray data reveals transcript patterns associated with somatic embryogenesis in soybean. Plant Physiol 132:118–136.

Tokuji Y, Kuriyama K. 2003. Involvement of gibberellin and cytokinin in the formation of embryogenic cell clumps in carrot (*Daucus carota*). J Plant Physiol 160:133–141.

Ware D, Jaiswal P, Ni J, Pan X, Chang K, others. 2002. Gramene: a resource for comparative grass genomics. Nucleic Acids Res 30:103–105.

Winkelmann T, Hohe A, Pueschel A, Schwenkel H. 2000. Somatic embryogenesis in *Cyclamen persicum* Mill. Curr Topics Plant Biol 2:51–62.

Winkelmann T, Hohe A, Schwenkel H. 1998. Establishing embryogenic suspension cultures in Cyclamen persicum ''Purple Flamed.'' Adv Hort Sci 12:25–30.

Winkelmann T, Meyer L, Serek M. 2004. Desiccation of somatic embryos of *Cyclamen persicum* Mill. J Hort Sci Biotechnol 79:479–483.

Zimmerman L. 1993. Somatic embryogenesis: a model for early development in higher plants. Plant Cell 5:1411–1423.