Supplementary Information

Identification of functional 2A sequences

Fusion constructs of the two fluorescent proteins eGFP and sTomato, a variant of the red fluorescent protein, were generated as schematically depicted in Figure S1. The gene fusions were placed either behind the *AOX*1 or *GAP* promoter by cloning them into the *P. pastoris* expression plasmids pPp_T4_S and pPp_T4_GAP_S, respectively. The genes of eGFP and sTomato were fused only via the 2A sequences without any further linker. Thereby, the stop codon of the first gene in the polycistronic construct was omitted as well as the start codon of the subsequent gene. To evaluate if differences can be observed when the start codon is present on the second coding sequence, an additional construct pPp_T4_S_eGFP_FMDV2A_sTomato_withATG was generated. In addition, a 6xHis-Tag was added to the N-terminus of the gene fusion to allow Western blot analysis of the resulting gene products. To examine whether the position of the genes in the polycistronic construct does affect the respective expression levels, two series of constructs were generated harbouring either the fusion eGFP_sTomato or sTomato_eGFP. The corresponding vector maps can be found in the Appendix.

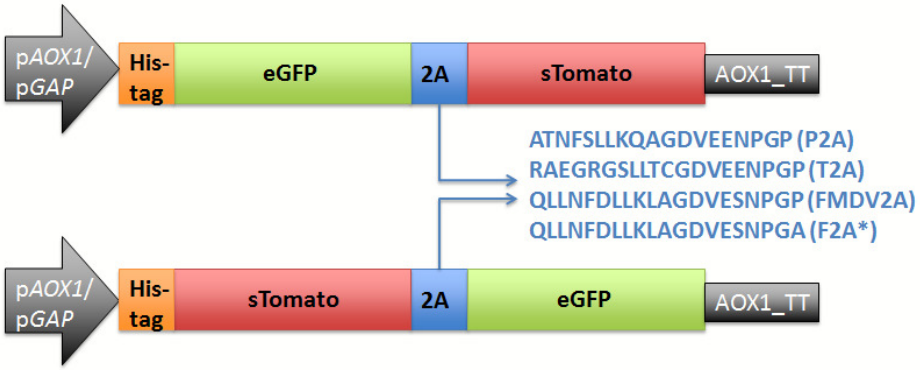

Figure S1: Schematic representation of the expression construct for testing the 2A activity *in P. pastoris*. The 2A sequence (marked in blue) was variable – four different sequences were tested (P2A, T2A, FMDV2A, F2A*).

The expression constructs were transformed into *P. pastoris* and the resulting transformants were screened for eGFP and sTomato fluorescence, respectively. In Figure S2 the screening results are exemplarily shown for the constructs pPp_T4_S_eGFP_P2A_sTomato (panel A and C) and pPp_T4_S_ sTomato_P2A_ eGFP. All tested transformants showed green (eGFP) and red (sTomato) fluorescence indicating that both proteins were functionally expressed. In addition, the corresponding fluorescence levels did not change significantly depending on the gene position in the polycistronic construct. However, the position on the single transcript might affect the expression levels in case that other more complex proteins are produced and/or more than two proteins are produced coordinately.

The construct pPp_T4_S_eGFP_FMDV2A_sTomato_withATG resulted in expression strains showing the same sTomato fluorescence as strains based on the construct where the corresponding ATG of the sTomato coding sequence was removed (data not shown). For similar results see also Figure S5.

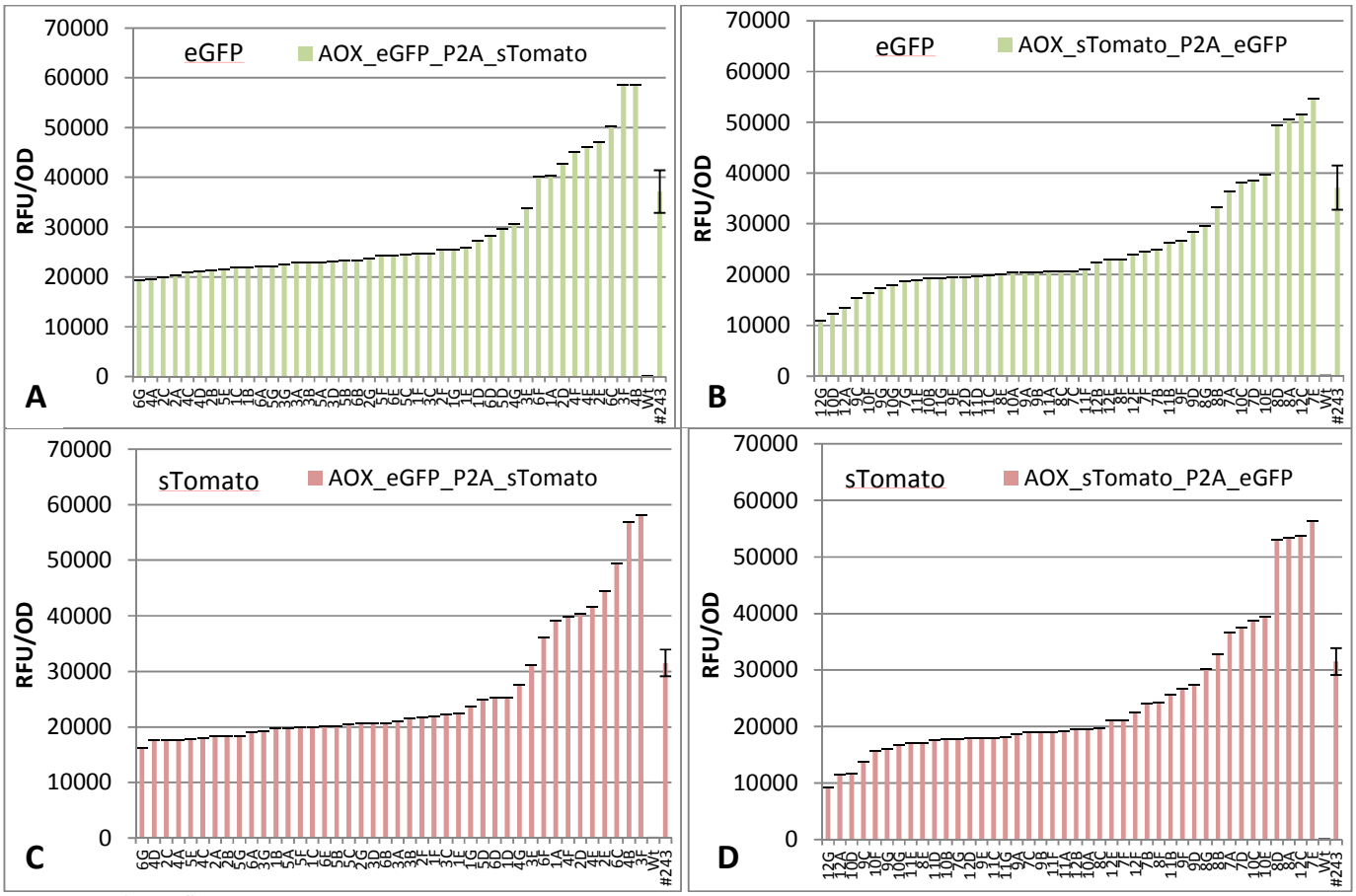

Figure S2: eGFP and sTomato fluorescence levels obtained by coordinate expression based on 2A sequences. Exemplarily, the screening results of the construct pPp_T4_S_eGFP_P2A_sTomato (panel A and C) and pPp_T4_S_ sTomato_P2A_ eGFP (panel B and D) are shown. *P. pastoris* CBS 7435 was used as negative control, strain #243 expressing eGFP and sTomato served as positive control.

Western blot analysis employing anti-bodies binding to the N-terminally attached His-tag was conducted to investigate whether the fluorescent proteins are present as fusions (~55 kDa) or as separate proteins (~27 kDa). In the case of P_{AOX1} driven expression, bands indicating the presence of separate fluorescence proteins as well as of the protein fusion were detected for all tested 2A sequences (Figure S3, upper panel). Also in the negative control eGFP-F2A-sTomato a band for His-tagged eGFP was observed. In this construct, the two fluorescent proteins were separated by a defective 2A sequence which contains a PGA instead of the PGP required for the ribosomal skip. The second negative control (sTomato-F2A-eGFP) only showed the expected band corresponding to the fusion product. These findings might indicate that the sequences surrounding the 2A sequences have an influence on the ribosoaml skipping mechanism too or represent a target for endogenous proteases causing protein cleavage.

In the case of P_{GAP} driven gene expression, only the constructs based on the T2A and P2A sequences resulted in separate fluorescence proteins (Figure S3, lower panel). Employing the FMDV2A as well as the defective F2A sequence only yielded the fusion product. It still needs to be clarified why the obtained results for P_{AOX1} and P_{GAP} are not the same. However, it was clearly shown that the 2A sequences from *Thosea asignus* virus and porcine teschovirus-1 are functional in *P. pastoris* allowing the coordinate expression of two genes and significant amounts of separate proteins.

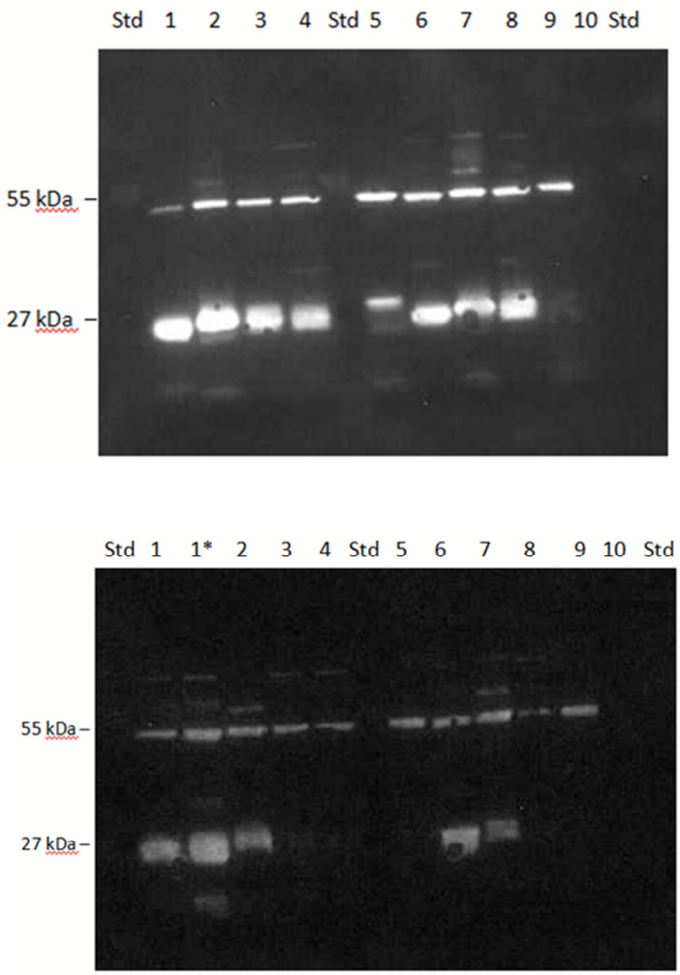

Figure S3: Western blot analysis of crude cell lysates of *P. pastoris* strains expressing different 2A constructs under the control of P_{AOX1} (upper panel) or P_{GAP} (lower panel) using anti-His antibody. The expected bands of the uncleaved protein fusion (55 kDa) and the single fluorescence proteins (27 kDa) are indicated. Lane 1: eGFP – T2A – sTomato; 1*: eGFP – T2A – sTomato, multicopy strain; 2: sTomato – T2A – eGFP; 3: eGFP – FMDV2A (with start codon) – sTomato; 4: eGFP – FMDV2A (without start codon) – sTomato; 5: sTomato – FMDV2A – eGFP; 6: eGFP – P2A – sTomato; 7: sTomato – P2A – eGFP; 8: eGFP – F2A – sTomato; 9: sTomato – F2A – eGFP; 10: *P. pastoris* CBS7435.

Polycistronic expression constructs in combination with an ubiquitin linker

The C- as well as the N-terminus of the proteins located up- and downstream of the 2A sequences is modified, respectively. In case that an authentic N-terminus is required a potential strategy is to add additional sequences that are post-translationally cleaved-off. Therefore, we tested an additional ubiquitin tag as autoprotease employing an expression construct as depicted in Figure S4. The two fluorescent proteins sTomato and eGFP (CDS with and without start codon) were fused via the T2A peptide with the ubiquitin tag intervening. The His-tag at the N-terminus of the expression construct was added to allow Western blot analysis of the resulting gene products.

Figure S4: Schematic representation of the polycistronic expression cassette coding for the two fluorescent proteins sTomato and eGFP carrying an additional ubiquitin linker.

The polycistronic expression construct with the additional ubiquitin linker did result in functional fluorescent proteins too. The expression levels of eGFP are in the same order of magnitude as the one obtained by a strain harboring the corresponding expression construct without ubiquitin (indicated as TTG in the landscapes of Figure S5). However, the expression of sTomato was affected negatively, as only about 50% of the red fluorescence was detected in comparison to the unmodified construct.

As already stated above, the obtained expression levels of the gene located downstream of the 2A sequence were in the same order of magnitude independent of the presence of an additional ATG of the corresponding coding sequence (compare Figure S5, panel A and B).

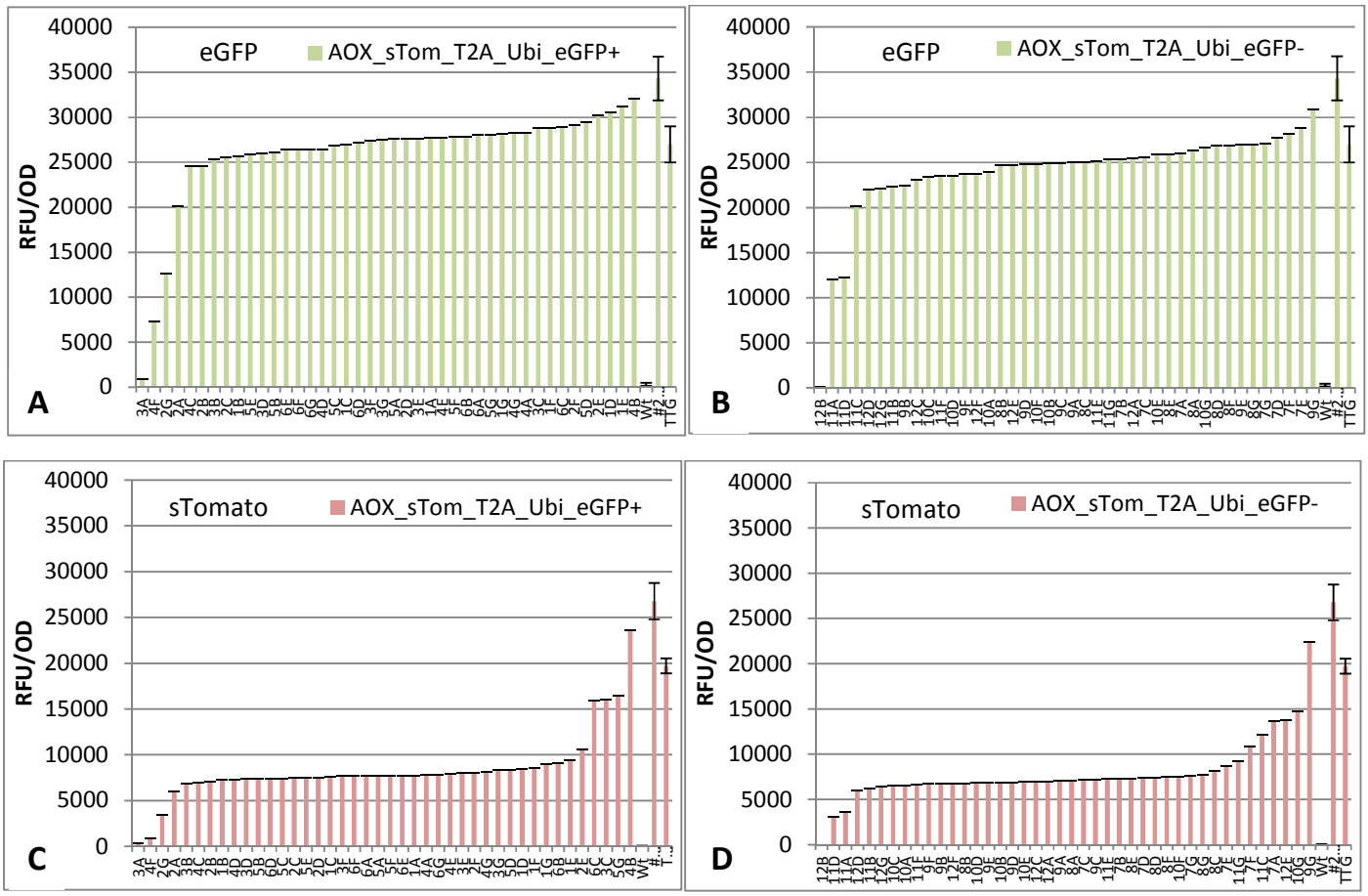

Figure S5: eGFP and sTomato fluorescence levels obtained by coordinate expression based on 2A sequences. The screening results of the construct pPp_T4_S_sTomato_T2A_Ubiquitin_eGFP with start codon of eGFP (panel A and C) and without start codon of eGFP (panel B and D) are shown. *P. pastoris* CBS 7435 was used as negative control, strain #243 expressing eGFP and sTomato served as positive control as well as the construct pPp_T4_S_sTomato_T2A_ eGFP (TTG, without ubiquitin).

Western blot analysis revealed that the additional ubiquitin linker did not interfere with the T2A-mediated cleavage (Figure S6). The predominant bands that were observed correspond to the cleaved His-tagged sTomato (~ 27kDa). A band corresponding to the full length gene fusion product (i.e. sTomato_eGFP) was only observed for the polycistronic expression construct without ubiquitin tag (Figure S6, lane 3). When ubiquitin was included, a smaller band was detected that would match the fluorescent protein sTomato carrying a C-terminal extension consisting of the 2A peptide and ubiquitin.

In addition, MS-analysis revealed that the resulting eGFP protein did not contain the 2A derived proline, but the natural N-terminus (methionine was not present independent of the presence of an ATG codon).

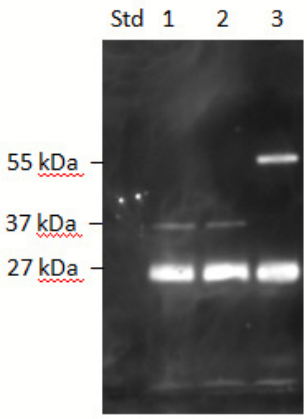

Figure S6: Western blot analysis of crude cell lysates of *P. pastoris* strains expressing the 2A-ubiquitin constructs under the control of P_{AOX1} using anti-His antibody. The expected bands of the uncleaved protein fusion (55k kDa for sTomato_T2A_ eGFP, 37 kDA for the product sTomato_T2A_Ubiquitin) and the fluorescence protein sTomato (27 kDa) are indicated. Lane 1: sTomato – T2A – ubiquitin – eGFP with start codon; 2: sTomato – T2A – ubiquitin – eGFP without start codon; 3: sTomato – T2A – eGFP 10C.

Material and Methods

General

Unless stated otherwise, all chemicals were obtained from Sigma-Aldrich (Steinheim, Germany) or Carl-Roth (Karlsruhe, Germany) with the highest purity available. Zeocin™ was obtained from InvivoGen (San Diego, CA, USA). Phusion® High Fidelity Polymerase for DNA amplification and further DNA modifying enzymes were purchased from Thermo Fisher Scientific Inc. (Waltham, MA, USA) or New England Biolabs (Ipswich, MA, USA). *E. coli* Top10 (Invitrogen, Carlsbad, USA) was used for all cloning steps and plasmid propagation. The *P. pastoris* strain CBS7435 as well as the plasmids pPp_T4_S and pPp_T4_GAP_S were obtained from the Pichia pool of TU Graz.[1] The wild type carotenoid pathway genes from *P. ananatis* and codon optimized versions of the violacein pathway genes from *C. violaceum* were obtained as synthetic genes from GeneScript (Piscataway, NJ, USA).

Plasmid and strain generation

All expression constructs generated during this study are summarized in Table S1. Overlap-extension PCR as well as Gibson cloning [2] were employed for expression construct assembly. Primer sequences, 2A sequences as well as the corresponding plasmid maps are provided in the appendix.

Table S1: Expression constructs assembled during the present study.

Expression construct	**Remarks**
pPp_T4_S_eGFP_T2A_sTomato	Construct also available with P_{GAP}
pPp_T4_S_ sTomato _T2A_ eGFP	Construct also available with P_{GAP}
pPp_T4_S_eGFP_P2A_sTomato	Construct also available with P_{GAP}
pPp_T4_S_ sTomato _P2A_ eGFP	Construct also available with P_{GAP}
pPp_T4_S_eGFP_F2A_sTomato	Construct also available with P_{GAP}
pPp_T4_S_ sTomato _F2A_ eGFP	Construct also available with P_{GAP}
pPp_T4_S_eGFP_FMDV2A_sTomato	Construct also available with P_{GAP}; in addition constructs harboring the CDS of sTomato including the start codon
pPp_T4_S_ sTomato _FMDV2A_ eGFP	Construct also available with P_{GAP}
pPp_T4_S_crtEBIY, T2A	Construct also available with P_{GAP}
pPp_T4_S_crtEBIY, T2A_P2A	Construct also available with P_{GAP}
pPp_T4_S_vioCBEDA, T2A	Construct also available with P_{GAP}
pPp_T4_S_crtEBIY_vioCBEDA, T2A	Construct also available with P_{GAP}
pPp_T4_S_vioABEDC_ crtEBIY_, T2A	Construct also available with P_{GAP}
pPp_T4_S_vioCBEDA_pHTX1_crtEBIY	Combination with bidirectional promoter (constitutive)
pPp_T4_S_vioCBEDA_pBZ6_crtEBIY	Combination with bidirectional promoter (inducible)
pPp_T4_S_sTomato_Ubiquitin_eGFP	Construct also available with P_{GAP}; in addition constructs harboring the CDS of eGFP including the start codon

All constructs were transformed as linear expression cassettes into *P. pastoris* CBS7435 according to the condensed protocol by Lin-Cereghino *et al.* [3] Transformants were selected on YPD agar plates containing 100 mg/L Zeocin™.

Cultivation of *P. pastoris* strains

Protein expression in *P. pastoris* was performed essentially as described in [4]. Therefore, *Pichia* cultures were grown in buffered minimal dextrose (BMD) or buffered mineral methanol (BMM) medium containing 200 mM KP_i, pH 6.0, 13.4 g/L yeast nitrogen base and 0.4 mg/L biotin supplemented with 2 % (w/v) glucose or 5 % (v/v) methanol, respectively.

Fluorescence measurements

For the measurement of fluorescence 190 µL ddH_2O were mixed with 10 µL of liquid cultures of *P. pastoris* strains. Fluorescence of eGFP (488 nm excitation, 507 nm emission) and sTomato (544 nm excitation, 581 nm emission) was recorded with a Synergy MX Microplate Reader.

SDS-PAGE/Immunoblot analysis

Protein isolation from yeast was performed with the Y-PER™ Yeast Protein Extraction Reagent from Thermo Scientific Inc. according to the manufacturer´s instructions. The total protein concentrations of the obtained samples were determined by the Bio-Rad protein assay (Bio-Rad Laboratories GmbH, Germany) using BSA as standard. 2 µg of total protein per lane were separated by SDS-PAGE under reducing conditions using NuPAGE® 4-12% Bis-Tris gel (Invitrogen). Protein bands were transferred onto a nitrocellulose membrane (GE Healthcare, Chalfont St Giles, UK) electrophoretically in a wet blotting system. Immunoblot detection was performed using a HIS-specific antibody (Tetra His-antibody from Quiagen) as primary antibody as well as Goat Anti-Mouse IgG (H+L) – HRP from Invitrogen as secondary antibody according to the manual provided by the supplier.

MS-analysis

Protein isolation from *Pichia* strains harboring the construct pPp_T4_S_sTomato_T2A_eGFP pPp_T4_S_sTomato_T2A_Ubiquitin_eGFP (with and without start codon of the eGFP CDS) and SDS-PAGE thereof was conducted as described above. The gel band corresponding to eGFP (~27 kDa) was excised and digested with Endoproteinase Asp-N (Sigma). Peptide extracts were dissolved in 0,1% formic acid, 5% acetonitril and separated by nano-HPLC (Dionex Ultimate 3000) equipped with a C18, 5 µm, 100 Å, 5 x 0.3 mm enrichment column and an Acclaim PepMap RSLC nanocolumn (C18, 2 µm, 100 Å, 500 x 0.075 mm) (all Thermo Fisher Scientific, Vienna, Austria). Samples were concentrated on the enrichment column for 2 min at a flow rate of 20 µL/min with 0.5 % trifluoroacetic acid as isocratic solvent. Separation was carried out on the nanocolumn at a flow rate of 200 nL/min using the following gradient, where solvent A is 0.3 % formic acid in water and solvent B is a mixture of 80% acetonitrile in water containing 0.3 % formic acid: 0-2 min 4 % B, 2-35 min 4-28% B, 35-47 min 28-50% B, 47-48 min 50-95% B, 48-58 min 95% B, 58-58,1 min 95-4% B, 58,1-70 min 4% B. The sample was ionized in the nanospray source equipped with stainless steel emitters (ES528, Thermo Fisher Scientific, Vienna, Austria) and analysed in a Orbitrap velos pro mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) operated in positive ion mode, applying alternating full scan MS (m/z 400 to 2000) in the ion cyclotron and MS/MS by higher-energy collisional dissociation of the 20 most intense peaks with dynamic exclusion enabled. The LC-MS/MS data were analysed by searching a database containing the protein sequences of eGFP and known background proteins with Mascot 2.3 (MatrixScience, London, UK). Detailed search criteria; enzyme: semispecific Asp-N, maximum missed cleavage sites: 2, N-terminus: hydrogen, C-terminus : free acid, Cys modification: carbamidomethylation, search mode: homology search, possible multiple oxidised methionine, maximum precursor charge 3; precursor mass tolerance +/- 0.05 Da, product mass tolerance +/- 0.7 Da., 5 % false discovery rate. Data was filtered according to stringent peptide acceptance criteria, including mass deviations of ±10 ppm, minimum 2 peptides per protein, Mascot Ion Score of at least 17 and a position rank 1 in Mascot search. For detection of possible

amino acid modifications or mutations, the LC-MS/MS data from the Asp-N digest was subjected to error tolerant search by Mascot 2.3 (MatrixScience, London, UK).

Product analysis

A small pellet of coloured *Pichia* cells was resuspended in 1 mL yeast lysis buffer (1 M sorbitol, 100 mM EDTA, 14 mM β-mercaptoethanol). 100 µL of a zymolyase stock solution (1000 U/mL) were added and the reaction mixture was incubated at 30°C for 30 min. The thus generated spheroplasts were pelleted by centrifugation (5 min, max. speed) and resuspended in 500 µL MeOH. Pigments were extracted by incubating the mixture twice for 15 min at 60°C. The combined organic phases were dried using a stream of dry nitrogen gas and dissolved in 100 µL MeOH. Extracts were subjected to TLC using an ethyl acetate/cyclohexane solvent system (9:1).

References

1. L. Näätsaari, B. Mistlberger, C. Ruth, T. Hajek, F. Hartner, and A. Glieder, *PLoS One*, 2012, **7**.

2. D. Gibson, L. Young, R. Chuang, J. Venter, C. Hutchison, and H. Smith, *Nat Methods*, 2009, **6**, 343–5.

3. J. Lin-Cereghino, W. W. Wong, S. Xiong, W. Giang, L. T. Luong, J. Vu, S. D. Johnson, and G. P. Lin-Cereghino, *Biotechniques*, 2005, **38**, 44, 46, 48.

4. R. Weis, R. Luiten, W. Skranc, H. Schwab, M. Wubbolts, and A. Glieder, *FEMS Yeast Res.*, 2004, **5**, 179–89.

Appendix

Nucleotide sequences of employed proteins

Fluorescent proteins

eGFP

ATGGCTAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGC
ACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCTACATACGGAAAGCTTACCCTTAAATTTATTTGCACTACT
GGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTGACTTATGGTGTTCAATGCTTTTCCCGTTATCCGGAT
CATATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAG
ATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAAGG
TATTGATTTTAAAGAAGATGGAAACATTCTCGGACACAAACTTGAGTACAACTATAACTCACACAATGTATACATC
ACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGTCACAACATTGAAGATGGTTCCGTTCAA
CTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGAC
ACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGG
ATTACACATGGCATGGATGAATTGTACAAGTAA

sTomato

ATGGTTTCTAAGGGTGAGGAAGTTATCAAGGAGTTCATGCGTTTCAAGGTCAGAATGGAAGGTTCTATGAACGGTC
ATGAGTTCGAGATTGAAGGAGAAGGTGAAGGAAGACCATATGAGGGTACTCAAACCGCAAAGTTGAAGGTTACTA
AAGGAGGTCCTTTACCATTCGCTTGGGATATCCTGTCTCCACAATTCATGTATGGTTCTAAGGCATACGTTAAGCAT
CCTGCAGACATTCCTGACTACAAGAAGTTGTCCTTTCCTGAGGGTTTCAAGTGGGAAAGAGTCATGAACTTCGAAG
ACGGTGGATTGGTGACTGTCACTCAAGACTCTTCCCTTCAAGACGGTACTTTGATCTACAAGGTCAAGATGCGTGGT
ACCAACTTCCCACCAGATGGTCCTGTTATGCAGAAAAAGACTATGGGATGGGAAGCTTCTACTGAGAGATTGTATC
CAAGAGATGGTGTTTTGAAGGGTGAGATTCACCAAGCTTTGAAGCTTAAAGATGGAGGTCACTACTTGGTTGAGTT
CAAGACCATTTACATGGCTAAGAAACCAGTTCAACTTCCTGGATACTATTACGTTGACACTAAGCTGGACATTACCT
CTCACAACGAAGACTACACCATCGTTGAGCAATACGAGAGATCCGAAGGTAGACACCACTTGTTCTTGTACGGTAT
GGACGAGCTTTATAAGTAA

Ubiquitin

ATGCAAATCTTCGTTAAGACTTTGACTGGTAAGACCATCACTCTTGAGGTTGAGTCTTCCGACACTATCGACAACGT
CAAGTCCAAGATCCAAGACAAGGAAGGTATTCCACCTGATCAGCAAAGACTGATCTTCGCTGGTAAACAGTTGGA
AGATGGTAGAACTTTGTCTGACTACAACATCCAGAAAGAGTCCACTTTGCACTTGGTCCTTAGACTTAGAGGAGGT

Carotenoid pathway

crtE

ATGACGGTCTGCGCAAAAAAACACGTTCATCTCACTCGCGATGCTGCGGAGCAGTTACTGGCTGATATTGATCGAC
GCCTTGATCAGTTATTGCCCGTGGAGGGAGAACGGGACGTTGTGGGTGCCGCGATGCGTGAAGGTGCGCTGGCACC
GGGAAAACGTATTCGCCCCATGTTGCTGTTGCTGACCGCCCGCGATCTGGGTTGCGCTGTCAGCCATGACGGATTA
CTGGATTTGGCCTGTGCGGTGGAAATGGTCCACGCGGCTTCGCTGATCCTTGACGATATGCCCTGCATGGACGATG
CGAAGCTGCGGCGCGGACGCCCTACCATTCATTCTCATTACGGAGAGCATGTGGCAATACTGGCGGCGGTTGCCTT
GCTGAGTAAAGCCTTTGGCGTAATTGCCGATGCAGATGGCCTCACGCCGCTGGCAAAAAATCGGGCGGTTTCTGAA
CTGTCAAACGCCATCGGCATGCAAGGATTGGTTCAGGGTCAGTTCAAGGATCTGTCTGAAGGGGATAAGCCGCGCA
GCGCTGAAGCTATTTTGATGACGAATCACTTTAAAACCAGCACGCTGTTTTGTGCCTCCATGCAGATGGCCTCGATT
GTTGCGAATGCCTCCAGCGAAGCGCGTGATTGCCTGCATCGTTTTTCACTTGATCTTGGTCAGGCATTTCAACTGCT
GGACGATTTGACCGATGGCATGACCGACACCGGTAAGGATAGCAATCAGGACGCCGGTAAATCGACGCTGGTCAA
TCTGTTAGGCCCGAGGGCGGTTGAAGAACGTCTGAGACAACATCTTCAGCTTGCCAGTGAGCATCTCTCTGCGGCC
TGCCAACACGGGCACGCCACTCAACATTTTATTCAGGCCTGGTTTGACAAAAAACTCGCTGCCGTCAGTTAA

crtB

ATGAATAATCCGTCGTTACTCAATCATGCGGTCGAAACGATGGCAGTTGGCTCGAAAAGTTTTGCGACAGCCTCAA
AGTTATTTGATGCAAAAACCCGGCGCAGCGTACTGATGCTCTACGCCTGGTGCCGCCATTGTGACGATGTTATTGAC
GATCAGACGCTGGGCTTTCAGGCCCGGCAGCCTGCCTTACAAACGCCCGAACAACGTCTGATGCAACTTGAGATGA
AAACGCGCCAGGCCTATGCAGGATCGCAGATGCACGAACCGGCGTTTGCGGCTTTTCAGGAAGTGGCTATGGCTCA
TGATATCGCCCCGGCTTACGCGTTTGATCATCTGGAAGGCTTCGCGATGGACGTACGCGAAGCGCAATACAGCCAA
CTGGACGATACGCTGCGCTATTGCTATCACGTTGCAGGCGTTGTCGGCTTGATGATGGCGCAAATCATGGGCGTGC
GGGATAACGCCACGCTGGACCGCGCCTGTGACCTTGGGCTGGCATTTCAGTTGACCAATATTGCTCGCGATATTGT
GGACGATGCGCATGCGGGCCGCTGTTATCTGCCGGCAAGCTGGCTGGAGCATGAAGGTCTGAACAAAGAGAATTA
TGCGGCACCTGAAAACCGTCAGGCGCTGAGCCGTATCGCCCGTCGTTTGGTGCAGGAAGCAGAACCTTACTATTTG
TCTGCCACAGCGGGCCTGGCAGGGTTGCCCCTGCGTTCCGCCTGGGCAATCGCTACGGCGAAGCAGGTTTACCGGA
AAATAGGTGTCAAAGTTGAACAGGCCGGTCAGCAAGCCTGGGATCAGCGGCAGTCAACGACCACGCCCGAAAAAT
TAACGCTGCTGCTGGCCGCCTCTGGTCAGGCCCTTACTTCCCGCATGCGGGCTCACCCTCCCCGCCCTGCGCATCTC
TGGCAGCGCCCGCTCTAG

crtI
ATGAAACCAACTACGGTAATTGGTGCAGGCTTCGGTGGCCTGGCACTGGCAATTCGTCTACAAGCTGCGGGGATTC
CCGTCTTACTGCTTGAACAACGTGATAAACCCGGCGGTCGGGCTTATGTCTACGAGGATCAGGGGTTTACCTTTGAT
GCAGGCCCGACGGTTATCACCGATCCCAGTGCCATTGAAGAACTGTTTGCACTGGCAGGAAAACAGTTAAAAGAGT
ATGTCGAACTGCTGCCGGTTACGCCGTTTTACCGCCTGTGTTGGGAGTCAGGGAAGGTCTTTAATTACGATAACGAT
CAAACCCGGCTCGAAGCGCAGATTCAGCAGTTTAATCCCCGCGATGTCGAAGGTTATCGTCAGTTTCTGGACTATTC
ACGCGCGGTGTTTAAAGAAGGCTATCTAAAGCTCGGTACTGTCCCTTTTTTATCGTTCAGAGACATGCTTCGCGCCG
CACCTCAACTGGCGAAACTGCAGGCATGGAGAAGCGTTTACAGTAAGGTTGCCAGTTACATCGAAGATGAACATCT
GCGCCAGGCGTTTTCTTTCCACTCGCTGTTGGTGGGCGGCAACCCCTTCGCCACCTCCTCCATTTATACGTTGATAC
ACGCGCTGGAGCGTGAGTGGGGCGTCTGGTTTCCGCGTGGCGGCACCGGCGCATTAGTTCAGGGTATGATAAAGCT
GTTTCAGGATCTGGGTGGCGAAGTCGTGTTAAACGCCAGAGTCAGCCACATGGAAACGACAGGAAACAAGATTGA
AGCCGTGCATTTAGAGGACGGTCGCAGGTTCCTGACGCAAGCCGTCGCGTCAAATGCAGATGTGGTTCATACCTAT
CGCGACCTGTTAAGCCAGCACCCTGCCGCGGTTAAGCAGTCCAACAAACTGCAGACTAAGCGCATGAGTAACTCTC
TGTTTGTGCTCTATTTTGGTTTGAATCACCATCATGATCAGCTCGCGCATCACACGGTTTGTTTCGGCCCGCGTTACC
GCGAGCTGATTGACGAAATTTTTAATCATGATGGCCTCGCAGAGGACTTCTCACTTTATCTGCACGCGCCCTGTGTC
ACGGATTCGTCACTGGCGCCTGAAGGTTGCGGCAGTTACTATGTGTTGGCGCCGGTGCCGCATTTAGGCACCGCGA
ACCTCGACTGGACGGTTGAGGGGCCAAAACTACGCGACCGTATTTTTGCGTACCTTGAGCAGCATTACATGCCTGG
CTTACGGAGTCAGCTGGTCACGCACCGTATGTTTACGCCGTTTGATTTTCGCGACCAGCTTAATGCCTATCATGGCT
CAGCCTTTTCTGTGGAGCCCGTTCTTACCCAGAGCGCCTGGTTTCGGCCGCATAACCGCGATAAAACCATTACTAAT
CTCTACCTGGTCGGCGCAGGCACGCACCCCGGCGCAGGCATTCCTGGCGTCATCGGCTCGGCAAAAGCGACAGCAG
GTTTGATGCTGGAGGATCTGATATGA

crtY
ATGCAACCGCATTATGATCTGATTCTCGTGGGGGCTGGACTCGCGAATGGCCTTATCGCCCTGCGTCTTCAGCAGCA
GCAACCTGATATGCGTATTTTGCTTATCGACGCCGCACCCCAGGCGGGCGGGAATCATACGTGGTCATTTCACCAC
GATGATTTGACTGAGAGCCAACATCGTTGGATAGCTCCGCTGGTGGTTCATCACTGGCCCGACTATCAGGTACGCTT
TCCCACACGCCGTCGTAAGCTGAACAGCGGCTACTTTTGTATTACTTCTCAGCGTTTCGCTGAGGTTTTACAGCGAC
AGTTTGGCCCGCACTTGTGGATGGATACCGCGGTCGCAGAGGTTAATGCGGAATCTGTTCGGTTGAAAAAGGGTCA
GGTTATCGGTGCCCGCGCGGTGATTGACGGGCGGGGTTATGCGGCAAATTCAGCACTGAGCGTGGGCTTCCAGGCG
TTTATTGGCCAGGAATGGCGATTGAGCCACCCGCATGGTTTATCTTCTCCCATTATCATGGACGCCACGGTCGATCA
GCAAAATGGTTATCGCTTCGTGTACAGCCTGCCGCTCTCGCCGACCAGATTGTTAATTGAAGACACGCACTATATTG
ATAATGCGACATTAGATCCTGAATGCGCGCGGCAAAATATTTGCGACTATGCCGCGCAACAGGGTTGGCAGCTTCA
GACACTGCTGCGAGAAGAACAGGGCGCCTTACCCATTACTCTGTCGGGCAATGCCGACGCATTCTGGCAGCAGCGC
CCCCTGGCCTGTAGTGGATTACGTGCCGGTCTGTTCCACCCTACCACCGGCTATTCACTGCCGCTGGCGGTTGCCGT
GGCCGACCGCCTGAGTGCACTTGATGTCTTTACGTCGGCCTCAATTCACCATGCCATTACGCATTTTGCCCGCGAGC
GCTGGCAGCAGCAGGGCTTTTTCCGCATGCTGAATCGCATGCTGTTTTTAGCCGGACCCGCCGATTCACGCTGGCGG
GTTATGCAGCGTTTTTATGGTTTACCTGAAGATTTAATTGCCCGTTTTTATGCGGGAAAACTCACGCTGACCGATCG
GCTACGTATTCTGAGCGGCAAGCCGCCTGTTCCGGTATTAGCAGCATTGCAAGCCATTATGACGACTCATCGTTAA

Violacein pathway
vioA
ATGAAACACTCTTCCGACATTTGTATTGTCGGAGCTGGTATCTCAGGTTTGACTTGCGCCTCTCACTTGCTGGATTCT
CCAGCCTGTAGAGGTTTGTCCCTTAGAATCTTTGACATGCAACAGGAGGCTGGAGGTAGAATTAGATCAAAGATGT
TGGATGGTAAGGCATCTATTGAATTGGGTGCTGGTAGATACTCTCCACAATTGCACCCACACTTTCAGTCTGCTATG
CAACATTACTCCCAAAAGTCTGAGGTTTACCCTTTCACTCAGCTTAAGTTCAAATCTCACGTTCAACAGAAGCTTAA
GCGTGCTATGAACGAATTGTCCCCACGTTTGAAGGAGCACGGTAAAGAGTCTTTCTTACAGTTCGTGTCCAGATAC
CAAGGACATGACTCTGCTGTTGGTATGATTAGATCAATGGGTTACGATGCATTGTTTCTTCCAGACATCTCTGCTGA
GATGGCATACGACATTGTTGGTAAGCATCCTGAGATTCAATCTGTTACCGATAACGACGCTAACCAGTGGTTCGCA
GCTGAGACTGGTTTTGCTGGATTGATTCAGGGAATCAAAGCTAAGGTCAAGGCAGCTGGTGCCAGATTCTCCCTTG
GTTACAGATTGCTGTCTGTGAGAACTGACGGTGATGGATATCTGTTGCAGCTTGCCGGTGACGACGGATGGAAGCT
TGAGCACCGTACTAGACACCTGATCCTTGCTATTCCACCTTCTGCAATGGCAGGTTTGAATGTTGACTTCCCAGAAG
CTTGGTCTGGAGCCAGATACGGTTCCTTGCCTTTGTTTAAGGGATTCTTAACCTACGGTGAACCTTGGTGGTTGGAC
TACAAGCTTGATGACCAAGTCCTGATCGTCGACAATCCATTGCGTAAAATCTACTTCAAGGGAGACAAGTACCTTT
TCTTCTACACTGACTCTGAGATGGCCAACTACTGGAGAGGTTGCGTCGCCGAAGGTGAGGATGGTTACTTAGAGCA
AATCCGTACCCACTTGGCTTCTGCTCTTGGAATCGTCAGAGAAAGAATCCCTCAACCTTTGGCCCACGTTCATAAGT
ACTGGGCACACGGTGTTGAGTTCTGTAGAGACTCAGATATCGACCACCCATCCGCATTGTCACATAGAGACTCTGG
TATCATTGCTTGTTCTGATGCCTATACCGAACACTGCGGATGGATGGAGGGAGGACTGTTGTCCGCCAGAGAAGCC
TCCCGTTTGCTTTTGCAAAGAATTGCTGCCTAA

vioB
ATGTCTATTTTGGACTTCCCAAGAATCCACTTTCGTGGTTGGGCAAGAGTTAACGCTCCAACCGCCAACAGAGATCC
ACACGGTCACATCGATATGGCCTCCAACACCGTCGCTATGGCCGGTGAACCATTCGACCTGGCCAGACATCCTACC
GAGTTCCACAGACACTTGAGATCCTTGGGTCCAAGATTTGGATTAGATGGAAGAGCTGACCCTGAAGGACCTTTCT

CACTGGCCGAGGGTTACAACGCCGCTGGAAACAATCACTTCTCTTGGGAGTCAGCTACTGTTTCTCACGTGCAATG
GGACGGTGGAGAAGCTGATAGAGGTGACGGATTGGTTGGTGCCAGACTGGCTCTTTGGGGTCATTACAATGACTAC
TTGAGAACCACTTTCAATAGAGCTAGATGGGTCGATTCCGACCCTACCAGAAGAGATGCTGCACAAATCTACGCCG
GACAGTTCACCATTTCTCCAGCAGGTGCAGGTCCAGGAACCCCTTGGTTGTTTACCGCCGACATTGACGATTCTCAC
GGTGCTCGTTGGACCCGTGGAGGTCACATTGCTGAAAGAGGAGGTCACTTCTTAGATGAAGAGTTCGGTCTTGCTA
GACTGTTTCAGTTCTCAGTTCCAAAAGACCACCCTCACTTCCTTTTCCACCCTGGTCCATTTGATTCTGAGGCTTGGA
GAAGACTGCAACTTGCTTTGGAGGACGATGACGTTTTGGGACTTACTGTTCAATACGCACTGTTCAACATGTCTACT
CCTCCTCAACCAAACTCTCCAGTGTTCCATGACATGGTCGGTGTTGTCGGTCTTTGGAGAAGAGGTGAGCTTGCATC
TTACCCAGCTGGTAGACTGTTGCGTCCAAGACAACCAGGTCTGGGTGACCTTACTTTGAGAGTCAACGGTGGTAGA
GTCGCTTTGAACTTAGCCTGTGCCATTCCATTCTCAACTAGAGCTGCTCAGCCTTCAGCACCTGATAGACTGACTCC
AGACCTGGGAGCAAAGTTGCCATTGGGTGATTTGTTGTTACGTGACGAAGACGGAGCCCTGTTGGCTAGAGTCCCA
CAAGCACTTTACCAAGATTACTGGACTAACCACGGTATCGTTGACTTGCCTCTGTTGCGTGAGCCAAGAGGTTCCCT
TACCTTGTCTTCCGAGTTGGCAGAGTGGAGAGAGCAGGACTGGGTGACTCAATCTGATGCATCCAACCTTTACTTA
GAGGCTCCAGACAGACGTCACGGTCGTTTCTTTCCAGAGTCTATTGCTTTGCGTTCTTACTTCAGAGGTGAAGCAAG
AGCCAGACCTGACATTCCACACCGTATTGAGGGAATGGGTCTTGTGGGAGTTGAGTCAAGACAGGATGGTGACGCT
GCCGAATGGCGTCTGACCGGTTTGCGTCCAGGTCCTGCTAGAATCGTTCTGGACGATGGTGCAGAAGCTATTCCTTT
GAGAGTCCTTCCAGACGACTGGGCCTTGGATGACGCTACTGTTGAAGAGGTTGACTACGCTTTCTTGTACCGTCACG
TTATGGCCTACTATGAGTTGGTGTACCCTTTCATGTCTGATAAAGTCTTTTCTCTGGCTGACCGTTGTAAGTGTGAAA
CTTACGCCCGTTTGATGTGGCAAATGTGCGACCCACAAAACAGAAACAAGTCCTACTACATGCCTTCCACTAGAGA
GTTGTCCGCTCCAAAGGCTAGATTGTTTCTTAAGTACTTGGCTCACGTTGAAGGTCAAGCTAGATTGCAAGCTCCAC
CTCCAGCTGGTCCTGCACGTATTGAGTCTAAAGCTCAATTGGCCGCAGAATTGCGTAAGGCTGTTGACCTGGAGCTT
TCCGTTATGCTTCAATACCTTTACGCTGCCTACTCTATCCCTAACTATGCACAGGGTCAGCAAAGAGTTAGAGACGG
TGCCTGGACTGCTGAGCAGTTACAATTGGCATGTGGTTCTGGAGATAGACGTAGAGACGGAGGTATCAGAGCTGCA
TTGCTTGAGATTGCTCACGAGGAAATGATTCACTATCTGGTTGTCAACAACTTGTTGATGGCTCTTGGTGAACCTTT
CTACGCCGGTGTTCCATTGATGGGTGAAGCCGCTAGACAAGCATTCGGTTTAGACACCGAGTTTGCCTTAGAACCTT
TCTCTGAGTCCACCCTTGCCCGTTTCGTTCGTCTGGAATGGCCTCACTTCATCCCAGCACCTGGTAAGTCCATTGCTG
ACTGTTATGCTGCCATCAGACAAGCTTTCTTGGACCTGCCAGATTTGTTTGGAGGTGAAGCAGGTAAGAGAGGAGG
TGAGCACCATCTTTTCTTGAACGAGTTGACCAATAGAGCCCACCCTGGATACCAATTGGAAGTCTTTGACAGAGAC
TCTGCTCTTTTCGGTATTGCTTTCGTTACTGACCAAGGTGAGGGTGGTGCTTTGGACTCTCCTCATTACGAACACTCC
CACTTTCAAAGATTGCGTGAGATGTCCGCTCGTATCATGGCTCAATCCGCTCCTTTCGAACCAGCTTTGCCTGCTCTT
CGTAACCCAGTCTTGGACGAATCCCCAGGATGCCAAAGAGTCGCAGACGGTAGAGCCCGTGCATTGATGGCCTTGT
ACCAGGGTGTTTACGAATTGATGTTCGCTATGATGGCTCAACATTTCGCTGTCAAGCCACTGGGTTCTTTGAGACGT
TCAAGACTTATGAACGCAGCTATTGACCTTATGACTGGTTTGCTTAGACCTTTGTCTTGCGCTTTGATGAATCTTCCA
TCTGGTATCGCTGGTAGAACTGCTGGTCCACCTCTGCCAGGTCCTGTTGACACTAGATCCTACGACGACTATGCTTT
AGGTTGTAGAATGTTGGCTAGAAGATGTGAGAGATTACTTGAGCAAGCCTCCATGTTGGAACCTGGTTGGCTTCCT
GATGCTCAGATGGAGTTGTTGGACTTCTACCGTAGACAGATGTTGGACTTGGCTTGTGGAAAGTTATCTCGTGAAG
CTTAA

vioC
ATGAAGAGAGCTATCATTGTTGGTGGAGGTTTAGCTGGTGGTTTGACCGCTATCTATCTGGCAAAGAGAGGTTACG
AAGTGCACGTTGTCGAGAAGAGAGGTGACCCACTTAGAGATTTGTCTTCCTACGTTGACGTCGTTTCCTCACGTGCC
ATTGGTGTGTCTATGACTGTTAGAGGTATTAAATCAGTCCTGGCTGCCGGTATTCCACGTGCTGAATTGGACGCTTG
TGGTGAGCCTATCGTTGCTATGGCCTTTTCTGTGGGTGGACAATACAGAATGAGAGAATTGAAACCTTTGGAAGAC
TTCCGTCCTCTTTCATTGAACAGAGCTGCATTTCAAAAGTTACTGAACAAGTATGCTAACTTGGCTGGTGTTAGATA
TTACTTCGAGCACAAGTGCCTTGATGTTGATCTGGACGGTAAGTCTGTCTTGATTCAGGGTAAGGACGGTCAACCA
CAAAGATTACAGGGTGATATGATCATTGGTGCCGATGGTGCACACTCAGCTGTGCGTCAAGCTATGCAATCTGGTT
TGAGACGTTTCGAGTTTCAACAGACCTTCTTCAGACACGGATACAAAACTTTGGTTCTTCCAGACGCTCAAGCCTTG
GGTTACAGAAAAGATACCTTGTACTTCTTTGGTATGGACTCCGGTGGATTGTTCGCAGGTAGAGCCGCTACTATTCC
AGACGGTTCTGTCTCTATCGCTGTTTGTCTTCCATACTCCGGATCTCCATCTTTGACTACTACCGACGAACCAACTAT
GCGTGCTTTCTTTGACAGATACTTCGGAGGTCTGCCAAGAGACGCTCGTGACGAGATGTTAAGACAGTTCCTGGCT
AAGCCTTCCAACGACTTGATCAATGTTCGTTCTTCCACTTTTCACTACAAGGGTAACGTCTTGCTGTTAGGAGACGC
CGCTCATGCTACTGCTCCTTTCCTTGGTCAAGGTATGAACATGGCATTGGAGGACGCAAGAACCTTCGTCGAATTGC
TTGACAGACATCAAGGAGATCAGGATAAGGCATTCCCTGAGTTCACTGAATTGAGAAAGGTTCAAGCCGATGCTAT
GCAGGATATGGCCAGAGCTAACTATGACGTCCTGTCCTGTTCTAATCCAATCTTCTTCATGAGAGCCAGATACACTC
GTTACATGCATTCCAAGTTTCCTGGTCTTTACCCACCTGATATGGCAGAGAAACTGTACTTCACCTCCGAACCATAC
GATCGTCTTCAACAGATTCAAAGAAAGCAGAACGTCTGGTACAAGATTGGTAGAGTTAACTAA

vioD
ATGAAGATCCTTGTGATTGGTGCAGGACCAGCCGGTTTGGTGTTTGCTTCTCAACTTAAGCAGGCTAGACCATTGTG
GGCTATTGACATCGTTGAGAAGAACGACGAACAAGAGGTGTTGGGATGGGGAGTCGTTCTTCCTGGTAGACCAGG
ACAACACCCTGCCAACCCACTGTCTTATCTTGATGCTCCTGAAAGATTGAATCCACAGTTCTTAGAGGACTTCAAGC
TTGTCCACCACAACGAACCTTCCTTGATGTCAACTGGTGTGCTGTTGTGCGGAGTTGAGCGTAGAGGTTTAGTCCAT
GCCTTGAGAGACAAATGTCGTTCTCAAGGTATCGCCATTAGATTTGAGTCTCCTTTGCTTGAGCACGGTGAACTTCC

ATTGGCTGACTACGATCTTGTTGTGTTGGCTAACGGTGTCAACCATAAGACCGCCCACTTCACTGAAGCTCTGGTTC
CACAGGTTGACTATGGTAGAAACAAGTACATTTGGTACGGAACCTCTCAGTTGTTTGACCAAATGAACTTGGTTTTC
AGAACCCACGGTAAGGACATCTTCATTGCCCACGCCTACAAGTACTCCGATACTATGTCAACCTTCATTGTCGAATG
CTCTGAGGAAACTTATGCAAGAGCTAGACTGGGTGAAATGTCCGAGGAGGCCTCTGCTGAGTACGTTGCTAAGGTT
TTCCAAGCCGAATTGGGTGGACACGGTCTGGTTTCTCAGCCTGGTCTTGGTTGGCGTAACTTCATGACCTTGTCCCA
CGATCGTTGTCATGATGGAAAGTTGGTTCTGTTAGGTGATGCACTTCAATCTGGTCACTTCTCAATCGGTCACGGAA
CCACTATGGCCGTGGTTGTCGCTCAGTTGCTGGTCAAAGCACTTTGTACTGAAGACGGTGTTCCTGCTGCTTTGAAA
CGTTTTGAGGAAAGAGCATTGCCACTGGTTCAACTGTTCAGAGGTCATGCTGACAATTCCCGTGTCTGGTTTGAAAC
TGTTGAAGAGAGAATGCACCTGTCCTCTGCTGAGTTCGTCCAATCCTTCGACGCTAGACGTAAGTCTCTTCCACCTA
TGCCAGAAGCTTTGGCACAAAACCTTAGATACGCCTTGCAAAGATAA

vioE
ATGGAAAACCGTGAGCCACCTTTGCTTCCAGCTAGATGGTCTTCCGCATACGTTTCATATTGGTCTCCAATGTTGCC
TGACGATCAGCTGACTTCCGGTTACTGTTGGTTTGACTACGAGAGAGACATTTGCCGTATTGATGGACTTTTCAATC
CTTGGTCAGAGAGAGACACCGGTTACAGATTGTGGATGTCTGAAGTTGGAAACGCCGCTTCCGGTAGAACTTGGAA
ACAGAAGGTTGCTTACGGTAGAGAACGTACCGCTTTAGGTGAGCAACTGTGTGAAAGACCACTTGATGACGAGACT
GGTCCATTCGCCGAGTTGTTCCTTCCTCGTGACGTTTTGAGAAGACTGGGTGCTAGACACATCGGTCGTAGAGTCGT
TTTGGGTAGAGAAGCCGATGGATGGAGATATCAAAGACCTGGAAAGGGTCCATCTACCCTTTACTTGGACGCAGCT
TCTGGTACTCCTCTTAGAATGGTGACTGGTGATGAGGCATCTCGTGCTTCCTTGAGAGACTTTCCAAACGTGTCTGA
GGCTGAAATCCCAGACGCTGTCTTCGCCGCTAAGAGATAA

2A Sequences

Table S2: Nucleotide sequences of the 2A sequences used in the present study.

Identifier	Nucleotide sequence
T2A1	AGA GCT GAG GGT AGA GGT TCT TTG CTT ACT TGC GGT GAC GTT GAG GAA AAC CCA GGT CCA
T2A2	CGT GCC GAA GGA CGT GGA TCC CTT TTG ACC TGC GGA GAT GTC GAA GAG AAT CCT GGA CCT
T2A3	AGA GCA GAA GGT CGT GGC TCA TTG CTG ACT TGT GGC GAC GTG GAG GAA AAT CCC GGA CCA
T2A4	CGT GCA GAG GGC CGT GGT TCC TTA CTT ACC TGC GGT GAT GTG GAA GAA AAT CCA GGA CCC
T2A5	CGT GCC GAG GGT AGG GGA TCA CTT CTT ACA TGT GGA GAC GTC GAG GAG AAC CCT GGT CCA
T2A6	AGA GCT GAA GGA AGG GGT TCC CTG TTA ACG TGT GGC GAT GTT GAA GAG AAC CCC GGT CCT
T2A7	AGG GCA GAA GGC AGA GGA TCT CTG TTG ACT TGT GGT GAT GTA GAG GAG AAT CCC GGC CCA
T2A8	AGG GCG GAG GGG AGA GGC TCT CTT TTA ACT TGT GGA GAT GTG GAA GAG AAC CCA GGC CCT
P2A	GCT ACT AAC TTC TCT TTG CTT AAG CAA GCT GGT GAC GTT GAG GAA AAC CCA GGT CCA
FMDV2A	CAA TTG CTT AAC TTC GAC TTA TTG AAG CTT GCT GGT GAC GTT GAG TCT AAC CCA GGT CCA

Polycistronic expression constructs for testing the functionality of 2A sequences in *P. pastoris*

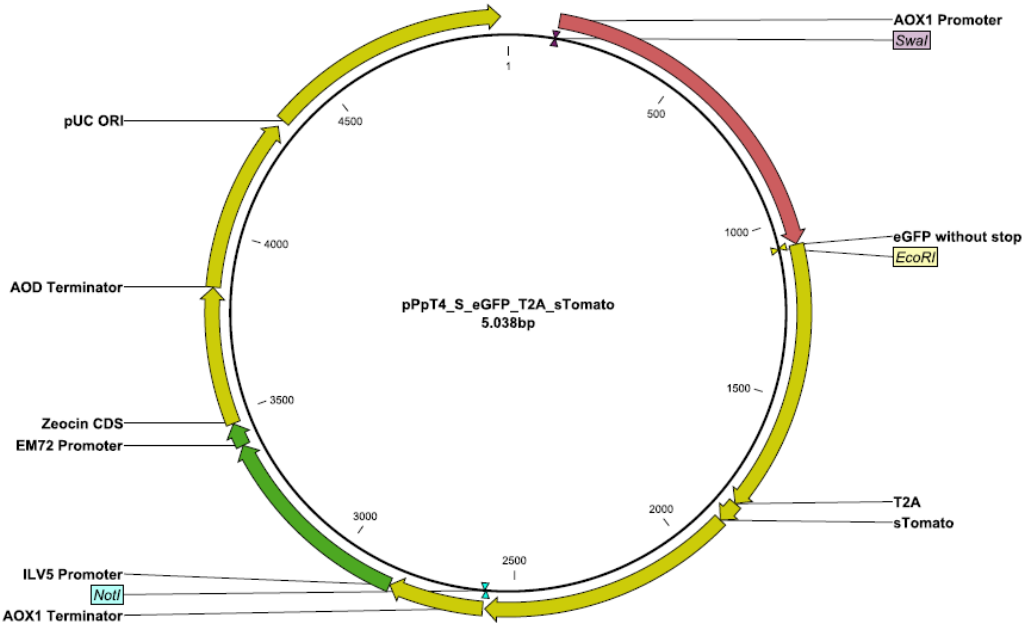

Figure S7: Exemplary plasmid map of the expression construct pPpT4_S_eGFP_T2A_sTomato. To test polycistronic gene expression in *P. pastoris* gene fusions of eGFP and sTomato and *vice versa* separated by diverse 2A sequences have been generated. The same set of constructs has also been generated based on the constitutive P_{GAP}.

Table S3: Primers used for the assembly of polycistronic expression constructs coding for eGFP and sTomato.

Name	**Sequence (5'→ 3')**
eGFP_EcoRI_fwd	AAA TGA ATT CCG AAA CGA TGG CTA GCA AAG GAG AAG AAC TTT TCA CTG
eGFP_FMDV2A_rev	TGG ACC TGG GTT AGA CTC AAC GTC ACC AGC AAG CTT CAA TAA GTC GAA GTT AAG CAA TTG CTT GTA CAA TTC ATC CAT GCC ATG TGT AAT CC
eGFP_P2A_rev	TGG ACC TGG GTT TTC CTC AAC GTC ACC AGC TTG CTT AAG CAA AGA GAA GTT AGT AGC CTT GTA CAA TTC ATC CAT GCC ATG TGT AAT CC
eGFP_T2A_rev	TGG ACC TGG GTT TTC CTC AAC GTC ACC GCA AGT AAG CAA AGA ACC TCT ACC CTC AGC TCT CTT GTA CAA TTC ATC CAT GCC ATG TGT AAT CC
sTomato_FMDV2A_fwd_mit Startcodon	CAA TTG CTT AAC TTC GAC TTA TTG AAG CTT GCT GGT GAC GTT GAG TCT AAC CCA GGT CCA ATG GTT TCT AAG GGT GAG GAA GTT ATC AAG GAG
sTomato_FMDV2A_fwd_ohne Startcodon	CAA TTG CTT AAC TTC GAC TTA TTG AAG CTT GCT GGT GAC GTT GAG TCT AAC CCA GGT CCA GTT TCT AAG GGT GAG GAA GTT ATC AAG GAG TTC ATG
sTomato_P2A_fwd	GCT ACT AAC TTC TCT TTG CTT AAG CAA GCT GGT GAC GTT GAG GAA AAC CCA GGT CCA GTT TCT AAG GGT GAG GAA GTT ATC AAG GAG TTC ATG
sTomato_T2A_fwd	AGA GCT GAG GGT AGA GGT TCT TTG CTT ACT TGC GGT GAC GTT GAG GAA AAC CCA GGT CCA GTT TCT AAG GGT GAG GAA GTT ATC AAG GAG TTC ATG
sTomato_NotI_rev	TAT TGC GGC CGC TTA CTT ATA AAG CTC GTC CAT ACC GTA CAA GAA CAA G
sTomato_EcoRI_fwd	AAA TGA ATT CCG AAA CGA TGG TTT CTA AGG GTG AGG AAG TTA TCA AGG AG
sTomato_FMDV2A_rev	TGG ACC TGG GTT AGA CTC AAC GTC ACC AGC AAG CTT CAA TAA GTC GAA GTT AAG CAA TTG CTT ATA AAG CTC GTC CAT ACC GTA CAA GAA CAA G
sTomato_P2A_rev	TGG ACC TGG GTT TTC CTC AAC GTC ACC AGC TTG CTT AAG CAA AGA GAA GTT AGT AGC CTT ATA AAG CTC GTC CAT ACC GTA CAA GAA CAA G
sTomato_T2A_rev	TGG ACC TGG GTT TTC CTC AAC GTC ACC GCA AGT AAG CAA AGA ACC TCT ACC CTC AGC TCT CTT ATA AAG CTC GTC CAT ACC GTA CAA GAA CAA G
eGFP_ FMDV2A_fwd	CAA TTG CTT AAC TTC GAC TTA TTG AAG CTT GCT GGT GAC GTT GAG TCT AAC CCA GGT CCA GCT AGC AAA GGA GAA GAA CTT TTC ACT GGA G
eGFP_P2A_fwd	GCT ACT AAC TTC TCT TTG CTT AAG CAA GCT GGT GAC GTT GAG GAA AAC CCA GGT CCA GCT AGC AAA GGA GAA GAA CTT TTC ACT GGA G
eGFP_T2A_fwd	AGA GCT GAG GGT AGA GGT TCT TTG CTT ACT TGC GGT GAC GTT GAG GAA

	AAC CCA GGT CCA GCT AGC AAA GGA GAA GAA CTT TTC ACT GGA G
eGFP_NotI_rev	TAT TGC GGC CGC TTA CTT GTA CAA TTC ATC CAT GCC ATG TGT AAT CC
sTomato_Gibson_rev	CTC TCA GGC AAA TGG CAT TCT GAC ATC CTC TTG AGC GGC CGC TTA CTT ATA AAG CTC GTC CAT ACC GTA CAA GAA CAA G
eGFP_Gibson_rev	CTC TCA GGC AAA TGG CAT TCT GAC ATC CTC TTG AGC GGC CGC TTA CTT GTA CAA TTC ATC CAT GCC ATG TGT AAT CC
eGFP_AOX_Gibson_His_fwd	ACG ACA ACT TGA GAA GAT CAA AAA ACA ACT AAT TAT TGA AAG AAT TCC GAA ACG ATG CAC CAC CAT CAC CAC CAT GCT AGC AAA GGA GAA GAA CTT TTC ACT G
sTomato_AOX_Gibson_His_fwd	ACG ACA ACT TGA GAA GAT CAA AAA ACA ACT AAT TAT TGA AAG AAT TCC GAA ACG ATG CAC CAC CAT CAC CAC CAT GTT TCT AAG GGT GAG GAA GTT ATC AAG GAG
eGFP_GAP_Gibson_His_fwd	GTC CCT ATT TCA ATC AAT TGA ACA ACT ATC AAA ACA CAG AAT TCC GAA ACG ATG CAC CAC CAT CAC CAC CAT GCT AGC AAA GGA GAA GAA CTT TTC ACT G
sTomato_GAP_Gibson_His_fwd	GTC CCT ATT TCA ATC AAT TGA ACA ACT ATC AAA ACA CAG AAT TCC GAA ACG ATG CAC CAC CAT CAC CAC CAT GTT TCT AAG GGT GAG GAA GTT ATC AAG GAG

Polycistronic expression of natural biosynthetic pathways

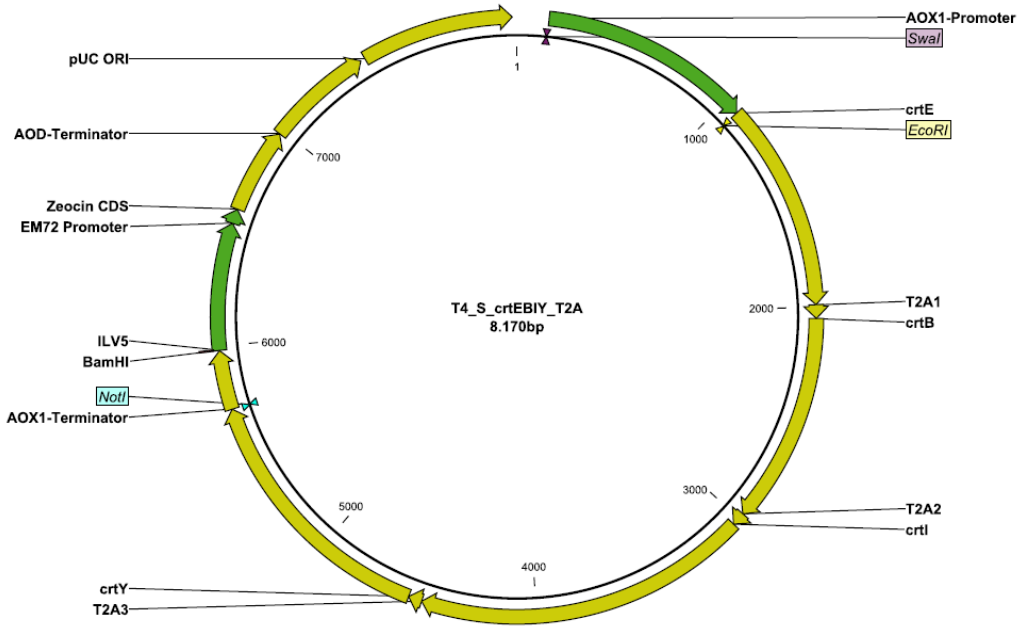

Figure S8: Plasmid map of the polycistronic expression construct coding for the carotenoid biosynthesis pathway based on P_{GAP}. The four pathway genes were fused via T2A sequences. An equivalent construct was generated based on the constitutive P_{GAP}.

Table S4: Primers used for the assembly of the polycistronic expression construct coding for the carotenoid biosynthesis pathway.

Name	Sequence (5'→ 3')
pAOX1_crtE_fw	CGA CAA CTT GAG AAG ATC AAA AAA CAA CTA ATT ATT GAA AGA ATT CCG AAA CGA TGA CGG TCT GC
crtE_T2A_rev	TGG ACC TGG GTT TTC CTC AAC GTC ACC GCA AGT AAG CAA AGA ACC TCT ACC CTC AGC TCT ACT GAC GGC AGC GAG TTT TTT GTC
crtB_T2A_fw	AGA GCT GAG GGT AGA GGT TCT TTG CTT ACT TGC GGT GAC GTT GAG GAA AAC CCA GGT CCA AAT AAT CCG TCG TTA CTC AAT CAT GCG G
crtB_T2A_rev	AGG TCC AGG ATT CTC TTC GAC ATC TCC GCA GGT CAA AAG GGA TCC ACG TCC TTC GGC ACGGAG CGG GCG CTG CCA GAG ATG
crtI_T2A_fw	CGT GCC GAA GGA CGT GGA TCC CTT TTG ACC TGC GGA GAT GTC GAA GAG AAT CCT GGA CCT AAA CCA ACT ACG GTA ATT GGT GCA GG
crtI_T2A_rev	TGG TCC GGG ATT TTC CTC CAC GTC GCC ACA AGT CAG CAA TGA GCC ACG ACC TTC TGC TCT TAT CAG ATC CTC CAG CAT CAA ACC TGC
crtY_T2A_fw	AGA GCA GAA GGT CGT GGC TCA TTG CTG ACT TGT GGC GAC GTG GAG GAA AAT CCC GGA CCA CAA CCG CAT TAT GAT CTG ATT CTC GTG G
crtY_AOX_TT_rev	CAG GCA AAT GGC ATT CTG ACA TCC TCT TGA GCG GCC GCT TAA CGA TGA GTC G
pGAP_crtE_fw	GTC CCT ATT TCA ATC AAT TGA ACA ACT ATC AAA ACA CAG AAT TCC GAA ACG ATG ACG GTC TGC
crtB_FMDV2A_rev	TGG ACC TGG GTT AGA CTC AAC GTC ACC AGC AAG CTT CAA TAA GTC GAA GTT AAG CAA TTG GAG CGG GCG CTG CCA GAG ATG
crtI_FMDV2A_fw	CAA TTG CTT AAC TTC GAC TTA TTG AAG CTT GCT GGT GAC GTT GAG TCT AAC CCA GGT CCA AAA CCA ACT ACG GTA ATT GGT GCA GG
crtI_P2A_rev	TGG ACC TGG GTT TTC CTC AAC GTC ACC AGC TTG CTT AAG CAA AGA GAA GTT AGT AGC TAT CAG ATC CTC CAG CAT CAA ACC TGC
crtY_P2A_fw	GCT ACT AAC TTC TCT TTG CTT AAG CAA GCT GGT GAC GTT GAG GAA AAC CCA GGT CCA CAA CCG CAT TAT GAT CTG ATT CTC GTG G

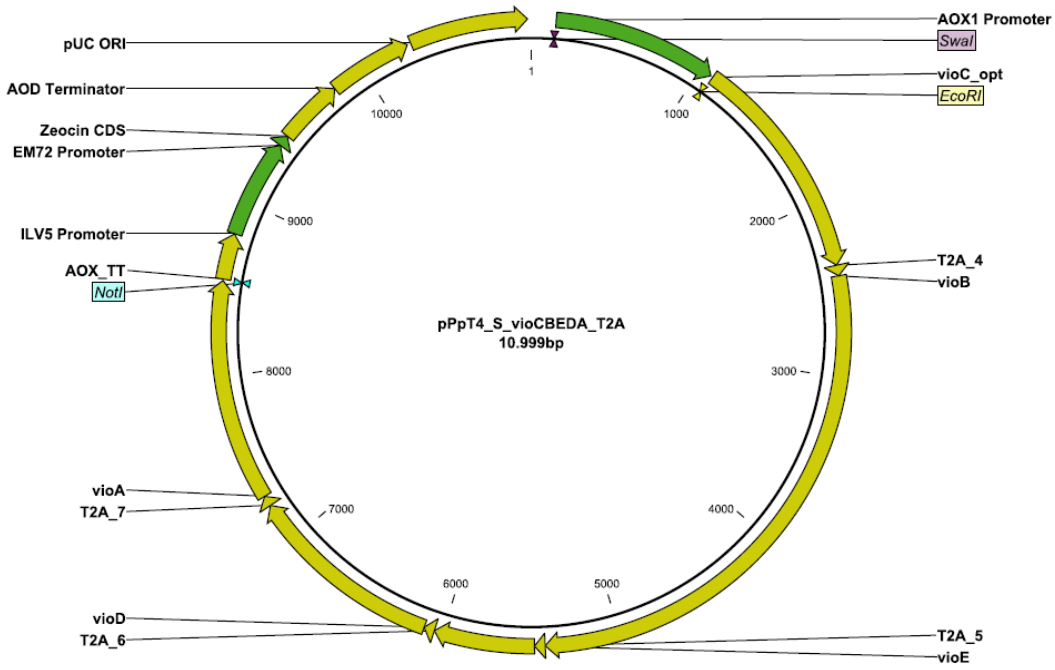

Figure S9: Plasmid map of the polycistronic expression construct coding for the violacein biosynthesis pathway based on P_{GAP}. The five pathway genes were fused via T2A sequences. An equivalent construct was generated based on the constitutive P_{GAP}.

Table S5: Primers used for the assembly of the polycistronic expression construct coding for the violacein biosynthesis pathway.

Name	**Sequence (5'→ 3')**
pGAP_vioC_fw	GTC CCT ATT TCA ATC AAT TGA ACA ACT ATC AAA ACA CAG AAT TCC GAA ACG ATG AAG AGA GCT ATC ATT G
pAOX1_vioC_fw	CGA CAA CTT GAG AAG ATC AAA AAA CAA CTA ATT ATT GAA AGA ATT CCG AAA CGA TGA AGA GAG CTA TCA TTG
vioC_T2A4_rev	GGG TCC TGG ATT TTC TTC CAC ATC ACC GCA GGT AAG TAA GGA ACC ACG GCC CTC TGC ACG GTT AAC TCT ACC AAT CTT GTA CCA GAC GTT C
T2A4_vioB_fw	CGT GCA GAG GGC CGT GGT TCC TTA CTT ACC TGC GGT GAT GTG GAA GAA AAT CCA GGA CCC TCT ATT TTG GAC TTC CCA AGA ATC CAC TTT C
vioB_T2A5_rev	TGG ACC AGG GTT CTC CTC GAC GTC TCC ACA TGT AAG AAG TGA TCC CCT ACC CTC GGC ACG AGC TTC ACG AGA TAA CTT TCC ACA AGC
T2A5_vioE_fw	CGT GCC GAG GGT AGG GGA TCA CTT CTT ACA TGT GGA GAC GTC GAG GAG AAC CCT GGT CCA GAA AAC CGT GAG CCA CCT TTG C
vioE_T2A6_rev	AGG ACC GGG GTT CTC TTC AAC ATC GCC ACA CGT TAA CAG GGA ACC CCT TCC TTC AGC TCT TCT CTT AGC GGC GAA GAC AGC G
T2A6_vioD_fw	AGA GCT GAA GGA AGG GGT TCC CTG TTA ACG TGT GGC GAT GTT GAA GAG AAC CCC GGT CCT AAG ATC CTT GTG ATT GGT GCA GGA C
vioD_T2A7_rev	TGG GCC GGG ATT CTC CTC TAC ATC ACC ACA AGT CAA CAG TGA TCC TCT GCC TTC TGC CCT TCT TTG CAA GGC GTA TCT AAG GTT TTG TG
T2A7_vioA_fw	AGG GCA GAA GGC AGA GGA TCT CTG TTG ACT TGT GGT GAT GTA GAG GAG AAT CCC GGC CCA AAA CAC TCT TCC GAC ATT TGT ATT GTC G
vioA_AOX_TT_rev	CAG GCA AAT GGC ATT CTG ACA TCC TCT TGA GCG GCC GCT TAG GCA GCA ATT CTT TGC AAA AGC AAA C

Polycistronic expression of a nine gene pathway

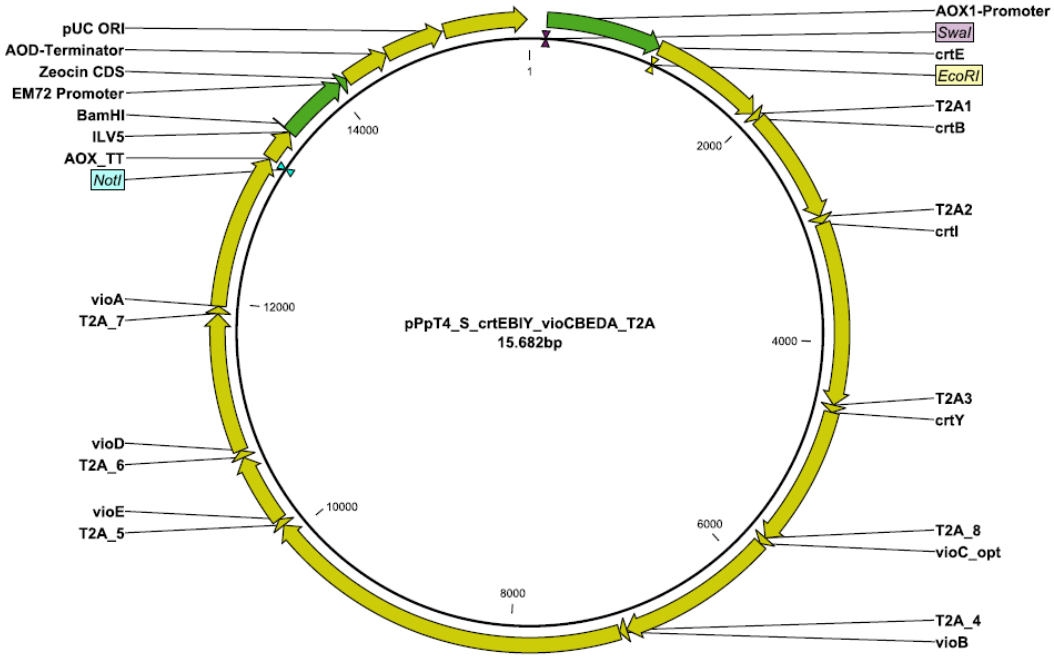

Figure S10: Plasmid map of the polycistronic expression construct coding for the carotenoid and the violacein biosynthesis pathway based on P_{GAP}. The nine pathway genes were fused via T2A sequences. An equivalent construct was generated based on the constitutive P_{GAP}.

Table S6: Primers used for the assembly of the polycistronic expression construct coding for the carotenoid and the violaceine biosynthesis pathway.

Name	**Sequence (5'→ 3')**
crtY_T2A8_rev	AGG GCC TGG GTT CTC TTC CAC ATC TCC ACA AGT TAA AAG AGA GCC TCT CCC CTC CGC CCT ACG ATG AGT CGT CAT AAT GGC TTG C
T2A8_ vioC_fw	AGG GCG GAG GGG AGA GGC TCT CTT TTA ACT TGT GGA GAT GTG GAA GAG AAC CCA GGC CCT AAG AGA GCT ATC ATT GTT GGT GGA GG
T2A_crtE_fw	AGA GCT GAG GGT AGA GGT TCT TTG CTT ACT TGC GGT GAC GTT GAG GAA AAC CCA GGT CCA GAA TTC CGA AAC GAT GAC GGT CTG C
crtB_T2A_rev	TGG ACC TGG GTT TTC CTC AAC GTC ACC GCA AGT AAG CAA AGA ACC TCT ACC CTC AGC TCT GAG CGG GCG CTG CCA GAG ATG
T2A_crtI_fw	AGA GCT GAG GGT AGA GGT TCT TTG CTT ACT TGC GGT GAC GTT GAG GAA AAC CCA GGT CCA AAA CCA ACT ACG GTA ATT GGT GCA GG
crtI_T2A_rev	TGG ACC TGG GTT TTC CTC AAC GTC ACC GCA AGT AAG CAA AGA ACC TCT ACC CTC AGC TCT TAT CAG ATC CTC CAG CAT CAA ACC TGC
T2A_crtY_fw	AGA GCT GAG GGT AGA GGT TCT TTG CTT ACT TGC GGT GAC GTT GAG GAA AAC CCA GGT CCA CAA CCG CAT TAT GAT CTG ATT CTC GTG G
crtY_T2A_rev	TGG ACC TGG GTT TTC CTC AAC GTC ACC GCA AGT AAG CAA AGA ACC TCT ACC CTC AGC TCT GCG GCC GCT TAA CGA TGA GTC G

Polycistronic expression combined with bidirectional promoters

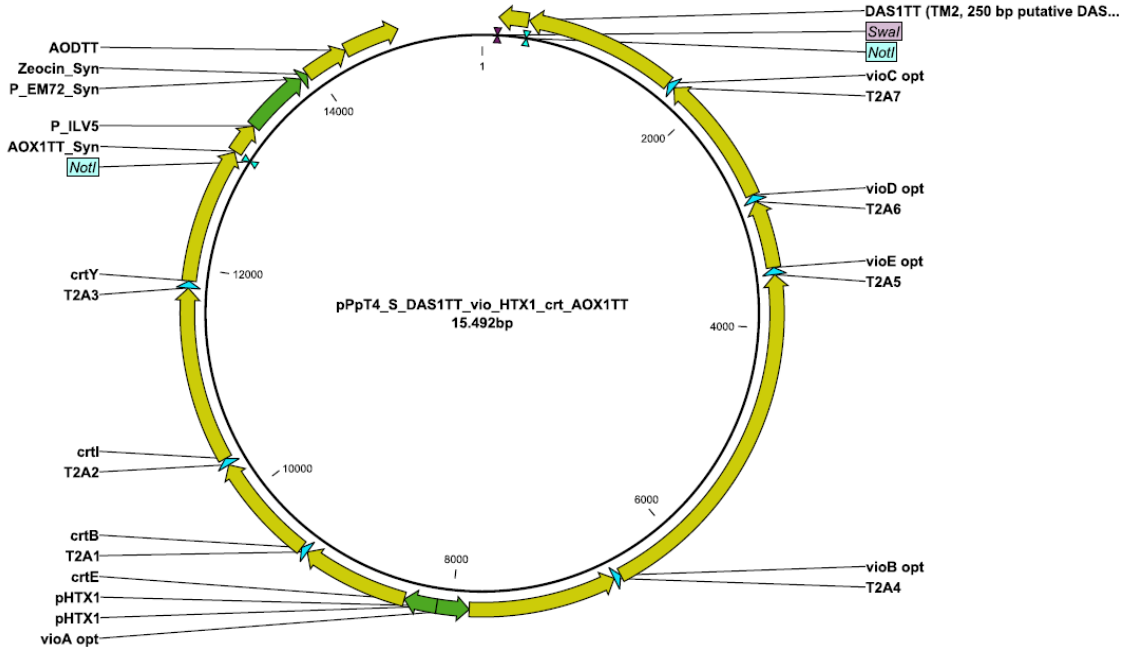

Figure S11: Plasmid map of the polycistronic expression construct coding for the carotenoid and the violacein biosynthesis pathway based on the bidirectional P_{HTX1}. The individual pathway genes were fused via T2A sequences. An equivalent construct was generated based on the inducible P_{BZ6}.

Table S7: Primers used for the assembly of the polycistronic expression construct coding for the carotenoid and the violaceine biosynthesis pathway in combination with a bidirectional promoter.

Name	**Sequence (5'→ 3')**
HTX_vioA_rev	CTC AAA CTA TAT TAA AAC TAC AAC AAT GAA ACA CTC TTC CGA CAT TTG TAT TGT CGG AG
vioA_HTX_fw	CTC CGA CAA TAC AAA TGT CGG AAG AGT GTT TCA TTG TTG TAG TTT TAA TAT AGT TTG AGT ATG AGA TGG AAC TCA
HTX_crtE_fw	CAA ACT AA TAC ATC CAG TTC AAG TTA CCT AAA CAA ATC AAA ATG ACG GTC TGC GCA AAA AAA CAC G
crtE_HTX_rev	CGT GTT TTT TTG CGC AGA CCG TCA TTT TGA TTT GTT TAG GTA ACT TGA ACT GGA TGT ATT AGT TTG
DAS1_TT_vioC_fw	GGG CTC CTA ACT AAA ACT GTA AAG ACT TCC CGT GCG GCC GCT TAG TTA ACT CTA CCA ATC TTG TAC CAG ACG TTC TGC
AOX1_TT_crtY_rev	GGC AAA TGG CAT TCT GAC ATC CTC TTG AGC GGC CGC TTA ACG ATG AGT CGT CAT AAT GGC TTG CAA TG
vioA_BZ6_fw	CGA CAA TAC AAA TGT CGG AAG AGT GTT TCA TTG TTG TAG TTT TAA TAT AGT TTG AGT ATG AGA TGG AAC TCAG
BZ6_vioA_rev	CTC AAA CTA TAT TAA AAC TAC AAC AAT GAA ACA CTC TTC CGA CAT TTG TAT TGT CG

Polycistronic expression constructs harboring ubiquitin as additional linker

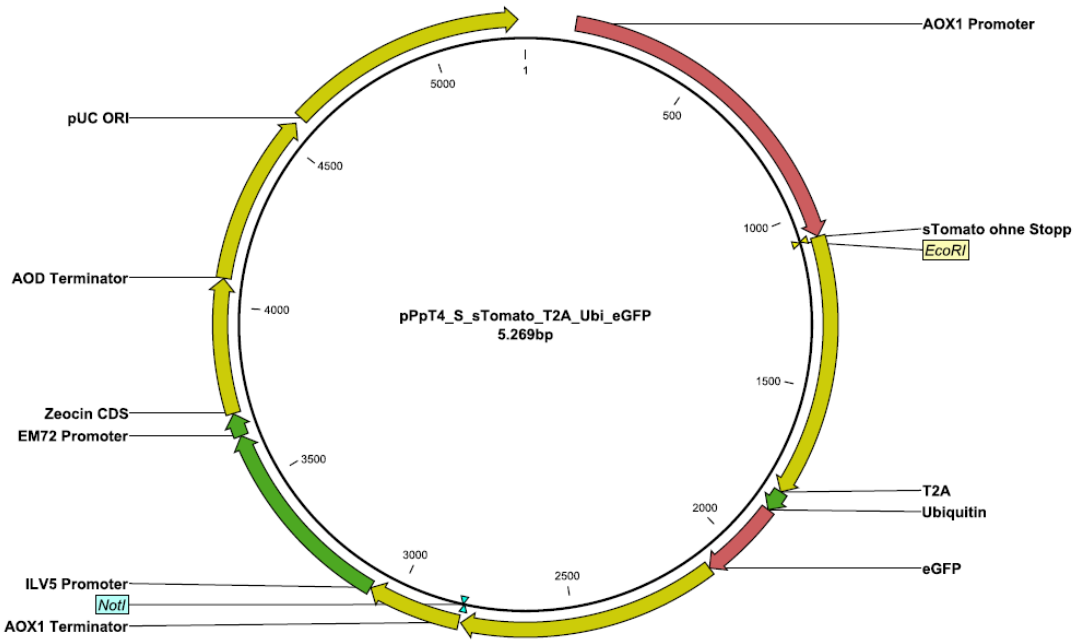

Figure S12: Plasmid map of the polycistronic expression construct coding for the fluorescent proteins sTomato and eGFP. The sequence of ubiquitin, an autoprotease, was located between the T2A sequence and the coding sequence of eGFP with and without start codon. An equivalent construct was generated based on the inducible constitutive P_{GAP}.

Table S8: Primers used for the assembly of the polycistronic expression construct harboring ubiquitin as additional linker.

Name	**Sequence (5'→ 3')**
T2A_Ubiqitin_GFP_fw	AGA GCT GAG GGT AGA GGT TCT TTG CTT AC
T2A_Ubiqitin_GFP_rev	GGG ACA ACT CCA GTG AAA AGT TCT TCT CC
Ubiquitin_GFP_fw	GCA CTT GGT CCT TAG ACT TAG AGG AGG TA TGG CTA GCA AAG GAG AAG AAC TTT TCA CTG
Ubiquitin_GFPohne_fw	GCA CTT GGT CCT TAG ACT TAG AGG AGG TG CTA GCA AAG GAG AAG AAC TTT TCA CTG
Ubiquitin_GFPohne_rev	CAG TGA AAA GTT CTT CTC CTT TGC TAG CAC CTC CTC TAA GTC TAA GGA CCA AGT GC
eGFP_Gibson_rev	CTC TCA GGC AAA TGG CAT TCT GAC ATC CTC TTG AGC GGC CGC TTA CTT GTA CAA TTC ATC CAT GCC ATG TGT AAT CC
sTomato_AOX_Gibson_His_fwd	ACG ACA ACT TGA GAA GAT CAA AAA ACA ACT AAT TAT TGA AAG AAT TCC GAA ACG ATG CAC CAC CAT CAC CAC CAT GTT TCT AAG GGT GAG GAA GTT ATC AAG GAG
sTomato_T2A_rev	TGG ACC TGG GTT TTC CTC AAC GTC ACC GCA AGT AAG CAA AGA ACC TCT ACC CTC AGC TCT CTT ATA AAG CTC GTC CAT ACC GTA CAA GAA CAA G