Supplementary Material

Nucleotide sequences of the codon-optimized brazzein gene and the three expression cassettes

This file provides the nucleotide sequences of the codon-optimized brazzein gene and of the three complete expression cassettes cloned into the pJAG vector backbone (BioGrammatics) for expression in *Pichia pastoris*. Each cassette encodes, in frame from the 5′ end: a secretion signal peptide, a 6×His tag, a TEV protease cleavage site (ENLYFQ↓G), and mature brazzein. Cleavage by TEV protease releases mature brazzein with an additional N-terminal glycine. The brazzein-coding portion is identical in all three constructs; the constructs differ only in the secretion signal.

Table S1 Overview of the three expression constructs used in this study

Plasmid	Signal	Cassette (bp)	Secretion signal peptide
pJAG-s1-bBRZN	s1	471	α-mating factor pre–pro signal (267 bp, 89 aa)
pJAG-s3-bBRZN	s3	480	Ost1p/α-mating factor fusion signal (276 bp, 92 aa)
pJAG-s5-bBRZN	s5	258	Aga2p signal (54 bp, 18 aa)

Table S2 Nucleotide sequences and conceptual translations of the codon-optimized brazzein gene, the purification and cleavage tags, the three secretion signals, and the three complete expression cassettes. a Codon-optimized brazzein coding sequence. b Purification and cleavage tags. c Secretion signal sequences. d Complete expression cassettes. The number at the left of each line indicates the position of the first nucleotide or residue on that line.

Table S2-a Codon-optimized brazzein coding sequence

The sequence below encodes mature brazzein preceded by the serine residue remaining after the Gly-Ser junction of the TEV site, followed by the stop codon (165 bp). This sequence is identical in pJAG-s1-bBRZN, pJAG-s3-bBRZN and pJAG-s5-bBRZN.

```
  1  TCCGACAAAT GTAAGAAGGT CTACGAGAAC TACCCAGTCT CCAAGTGCCA GTTGGCCAAC
 61  CAATGCAATT ATGATTGTAA ACTTGACAAG CACGCTCGTT CTGGTGAATG TTTCTACGAT
121  GAGAAGAGAA ACTTGCAATG TATCTGTGAC TACTGCGAGT ACTAG
```

Conceptual translation (54 aa plus stop):

```
1  SDKCKKVYEN YPVSKCQLAN QCNYDCKLDK HARSGECFYD EKRNLQCICD YCEY
```

Table S2-b Purification and cleavage tags

6×His tag (18 bp; HHHHHH):

```
1  CACCATCATC ATCACCAC
```

TEV protease cleavage site (21 bp; ENLYFQG):

```
1  GAAAATTTGT ACTTCCAAGG T
```

Table S2-c Secretion signal sequences

s1 — α-mating factor pre–pro secretion signal (267 bp):

```
    1  ATGAGATTCC CATCAATTTT TACTGCCGTT TTATTTGCTG CTTCTTCTGC TTTGGCCGCT
   61  CCGGTAAACA CCACTACTGA GGACGAAACT GCTCAAATTC CAGCTGAGGC TGTCATTGGT
  121  TACTTGGATT TGGAAGGAGA CTTCGATGTT GCCGTTCTTC CTTTCTCCAA CTCTACCAAC
  181  AACGGCTTGC TCTTTATCAA CACCACTATC GCTTCCATTG CTGCCAAGGA GGAAGGTGTT
  241  TCTTTGGACA AGAGAGAAGC TGAAGCT
```

```
Translation:
MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYLDLEGDFDVAVLPFSNSTNNGLLFINTTIASIAAKEEGVSLDKREAE
A
```

s3 — Ost1p/α-mating factor fusion secretion signal (276 bp):

```
    1  ATGAGACAAG TTTGGTTCTC ATGGATTGTT GGTTTGTTCC TGTGTTTCTT CAACGTCTCT
   61  TCTGCTGCTC CGGTAAACAC CACTACTGAG GACGAAACTG CTCAAATTCC AGCTGAGGCT
  121  GTCATTGGTT ACTTGGATTT GGAAGGAGAC TTCGATGTTG CCGTTCTTCC TTTCTCCAAC
  181  TCTACCAACA ACGGCTTGCT CTTTATCAAC ACCACTATCG CTTCCATTGC TGCCAAGGAG
  241  GAAGGTGTTT CTTTGGACAA GAGAGAAGCT GAAGCT
```

```
Translation:
MRQVWFSWIVGLFLCFFNVSSAAPVNTTTEDETAQIPAEAVIGYLDLEGDFDVAVLPFSNSTNNGLLFINTTIASIAAKEEGVSLDKR
EAEA
```

s5 — Aga2p secretion signal (54 bp):

```
    1  ATGCAATTGC TTAGATGCTT CTCTATTTTC TCCGTAATTG CATCTGTCTT GGCT
```

```
Translation: MQLLRCFSIFSVIASVLA
```

Table S2-d Complete expression cassettes

pJAG-s1-bBRZN — open reading frame, 471 bp (156 codons plus stop)

```
    1  ATGAGATTCC CATCAATTTT TACTGCCGTT TTATTTGCTG CTTCTTCTGC TTTGGCCGCT
   61  CCGGTAAACA CCACTACTGA GGACGAAACT GCTCAAATTC CAGCTGAGGC TGTCATTGGT
  121  TACTTGGATT TGGAAGGAGA CTTCGATGTT GCCGTTCTTC CTTTCTCCAA CTCTACCAAC
  181  AACGGCTTGC TCTTTATCAA CACCACTATC GCTTCCATTG CTGCCAAGGA GGAAGGTGTT
  241  TCTTTGGACA AGAGAGAAGC TGAAGCTCAC CATCATCATC ACCACGAAAA TTTGTACTTC
  301  CAAGGTTCCG ACAAATGTAA GAAGGTCTAC GAGAACTACC CAGTCTCCAA GTGCCAGTTG
  361  GCCAACCAAT GCAATTATGA TTGTAAACTT GACAAGCACG CTCGTTCTGG TGAATGTTTC
  421  TACGATGAGA AGAGAAACTT GCAATGTATC TGTGACTACT GCGAGTACTA G
```

Conceptual translation:

```
    1  MRFPSIFTAV LFAASSALAA PVNTTTEDET AQIPAEAVIG YLDLEGDFDV AVLPFSNSTN
   61  NGLLFINTTI ASIAAKEEGV SLDKREAEAH HHHHHENLYF QGSDKCKKVY ENYPVSKCQL
  121  ANQCNYDCKL DKHARSGECF YDEKRNLQCI CDYCEY
```

pJAG-s3-bBRZN — open reading frame, 480 bp (159 codons plus stop)

```
    1  ATGAGACAAG TTTGGTTCTC ATGGATTGTT GGTTTGTTCC TGTGTTTCTT CAACGTCTCT
   61  TCTGCTGCTC CGGTAAACAC CACTACTGAG GACGAAACTG CTCAAATTCC AGCTGAGGCT
  121  GTCATTGGTT ACTTGGATTT GGAAGGAGAC TTCGATGTTG CCGTTCTTCC TTTCTCCAAC
  181  TCTACCAACA ACGGCTTGCT CTTTATCAAC ACCACTATCG CTTCCATTGC TGCCAAGGAG
  241  GAAGGTGTTT CTTTGGACAA GAGAGAAGCT GAAGCTCACC ATCATCATCA CCACGAAAAT
  301  TTGTACTTCC AAGGTTCCGA CAAATGTAAG AAGGTCTACG AGAACTACCC AGTCTCCAAG
  361  TGCCAGTTGG CCAACCAATG CAATTATGAT TGTAAACTTG ACAAGCACGC TCGTTCTGGT
```

```
421  GAATGTTTCT ACGATGAGAA GAGAAACTTG CAATGTATCT GTGACTACTG CGAGTACTAG
```

Conceptual translation:

```
  1  MRQVWFSWIV GLFLCFFNVS SAAPVNTTTE DETAQIPAEA VIGYLDLEGD FDVAVLPFSN
 61  STNNGLLFIN TTIASIAAKE EGVSLDKREA EAHHHHHHEN LYFQGSDKCK KVYENYPVSK
121  CQLANQCNYD CKLDKHARSG ECFYDEKRNL QCICDYCEY
```

pJAG-s5-bBRZN — open reading frame, 258 bp (85 codons plus stop)

```
  1  ATGCAATTGC TTAGATGCTT CTCTATTTTC TCCGTAATTG CATCTGTCTT GGCTCACCAT
 61  CATCATCACC ACGAAAATTT GTACTTCCAA GGTTCCGACA AATGTAAGAA GGTCTACGAG
121  AACTACCCAG TCTCCAAGTG CCAGTTGGCC AACCAATGCA ATTATGATTG TAAACTTGAC
181  AAGCACGCTC GTTCTGGTGA ATGTTTCTAC GATGAGAAGA GAAACTTGCA ATGTATCTGT
241  GACTACTGCG AGTACTAG
```

Conceptual translation:

```
  1  MQLLRCFSIF SVIASVLAHH HHHHENLYFQ GSDKCKKVYE NYPVSKCQLA NQCNYDCKLD
 61  KHARSGECFY DEKRNLQCIC DYCEY
```

Notes

All open reading frames are placed under the control of the methanol-inducible AOX1 promoter and are followed by the AOX1 transcription terminator in the pJAG backbone. Vectors were linearized within the AOX1 promoter region prior to transformation to direct integration at the AOX1 locus of *P. pastoris* BG10.

Sequences were extracted from the verified plasmid maps and confirmed by conceptual translation against the protein sequence reported in the Materials and methods section (HHHHHHENLYFQGSDKCKKVYENYPVSKCQLANQCNYDCKLDKHARSGECFYDEKRNLQCICDYCEY).

A plain-text FASTA file containing the same sequences (Supplementary_Sequences.fasta) accompanies this document.