SUPPLEMENTARY FILE A: METHODS

1. Genome assemblies and annotation

All of the genome assemblies used in this study are publicly available from either the National Centre for Biotechnology Information (NCBI) or the Joint Genome Institute (JGI) (Table S1). All 14 genomes were subjected to gene annotation using the web-based gene annotator AUGUSTUS (Stanke and Waack, 2003). For the *Neurospora* genomes, the *N. crassa* species-specific gene model parameters were used. The *F. graminearum* parameters were used for all three *Huntiella* genomes, while the *Byssochlamys, Thermoascus* and *A. flavus* genomes were annotated using *A. nidulans* specific parameters. These species-specific parameters were chosen based on the closest relative that could be found in the AUGUSTUS gene model parameter library. In addition to these genome-wide annotations, the contigs on which the target genes (*MAT*, pheromone and pheromone receptor genes) were identified were also subjected to Fgenesh annotation (Salamov and Solovyev, 2000). The RNA sequence mapping described below was also used to confirm or correct the predicted gene models.

2. RNA-seq data

The majority of the RNA sequence data used in this study were obtained as raw reads from the Sequence Read Archive (SRA) and Gene Expression omnibus (GEO) databases of the NCBI (Table A.1).

Table A.1: The species used in this study, their sexual strategies, the genome assemblies used and, where relevant, the raw RNA reads used.

Species	Sexual Strategy	Isolate	Source	Accession Numbers	Reference / Submitter
Neurospora crassa	Heterothallic	OR7A FGSC4200 / 2489	NCBI	*DNA:* AABX00000000.3 *RNA:* GSE41484	Galagan et al., 2003 Wang et al., 2014
Neurospora discreta	Heterothallic	FGSC 8579	NCBI	RANX00000000.1	University of California, Berkeley
Neurospora tetrasperma	Pseudohomothallic	FGSC 2508	NCBI	AFBT00000000.1	Ellison et al., 2011
Neurospora terricola	Primary homothallic	FGSC 1889	NCBI	CAPR000000000.2	Gioti et al., 2013
Neurospora pannonica	Primary homothallic	FGSC 7221	NCBI	CAPQ000000000.2	Gioti et al., 2013
Neurospora africana	Unisexual (MAT1-1)	FGSC 1740	NCBI	*DNA:* CAPO000000000.2 *RNA:* ERS213531	Gioti et al., 2013 European Bioinformatics Institute
Huntiella bhutanensis	Heterothallic	CMW 8217	NCBI	MJMS00000000.1	Wingfield et al., 2016
Huntiella omanensis	Heterothallic	CMW 11056	NCBI	*DNA:* JSUI00000000.1 *RNA:* SRP108437	van der Nest et al., 2014 Wilson et al., 2018
Huntiella moniliformis	Unisexual (MAT1-2)	CMW 10134 CMW 36919	NCBI	*DNA:* JMSH00000000.1 *RNA:* SRP108437	Van Der Nest et al., 2014 Wilson et al., 2018
Byssochlamys spectabilis	Heterothallic	NBRC 109023	NCBI	RCHW00000000.1 SRP138004	Prefecture, 2014 Joint Genomes Institute
Byssochlamys fulva	Primary homothallic	BYSS01	NCBI	NIXA00000000.1	University of Dayton Research Institute
Byssochlamys nivea	Unisexual (MAT1-2)	CO 7	NCBI	QEIL00000000.1	Cornell University
Thermoascus aurantiacus	Unisexual (MAT1-1)	ATCC 26904	JGI NBCI	*DNA:* 405671 *RNA:* PRJNA681677	Singer, n.d. This study
Thermoascus crustaceus	Unisexual (MAT1-2)	JCM 12817	NGBI	BCIC00000000.1	RIKEN Center for Life Science Technologies
Aspergillus flavus	Primary homothallic	NRRL 3357	NCBI	AAIH00000000.2	Yin et al., 2018

2.1 RNA-seq data for Thermoascus aurantiacus

T. aurantiacus ATCC 26904 was grown in a modified glucose-preculture medium as previously described (Schuerg et al., 2017). Accordingly, ascospores from 7-day-old potato dextrose agar plates were harvested and 1 x 10^6 spores per ml were used to inoculate glucose pre-cultures (2% glucose [w/v], McClendon, pH = 5.25, 0.8% soy meal peptone [w/v]). 50 mL pre-cultures were incubated for 48 h in 250 mL Erlenmeyer flasks at 50 °C and 180 rpm. These cultures were filtered under vacuum, using a Buchner funnel and Whatman paper #1, each washed with 100 mL defined McClendon's medium without peptone (McClendon's salts pH = 5.25, 0.25 mM ammonia nitrate) (Mcclendon et al., 2012) and shifted to four new flasks containing the same medium and either 2% beechwood xylan (Megazyme) or no carbon source. All cultures were grown under constant incandescent light for 48 hours in a rotary shaker at 50 °C and 180 rpm. From these shift cultures, samples were taken 1, 4, 6 and 8 hours after shifting. The sampling procedure involved taking an aliquot of 8 ml culture and vacuum filtering and washing it on a Whatman paper with immediate freezing in liquid nitrogen.

Total nucleic acids were extracted by grinding the frozen culture samples with a mortar and pestle in liquid nitrogen and resuspending this in TRIzol® Reagent (Thermo Fisher Scientific). RNA was isolated with Phenol:Chloroform:Isoamyl Alcohol, ethanol precipitation and subsequent DNAseI treatment. RNA quantity was determined on a Qubit (Thermo Fisher Scientific) and RIN values were determined using a Bioanalyzer (Agilent RNA 6000 Nano). Samples of suitable quality (RIN values > 8) from all time points of xylan and no carbon cultures were taken for cDNA synthesis and library

construction with the TruSeq® Stranded Total RNA Library Kit (Illumina®). For this study, only one biological replicate for each condition and time point was sequenced on a MiSeq (Illumina®).

3. Gene identification

3.1 MAT genes

MAT genes for which reference sequences were available were downloaded from NCBI (Table A.2) and used in BLASTn queries against their respective genomes in order to identify the *MAT* loci. For example, the sequences for *matA-1* (*MAT1-1-1*), *matA-2* (*MAT1-1-2*) and *matA-3* (*MAT1-1-3*) from *N. crassa* were downloaded and used as a BLASTn query against the *N. crassa* genome. In order to identify the *B. spectabilis MAT* locus, the MAT1-1-1 protein sequence was downloaded from NCBI and used as a tBLASTn query against the *B. spectabilis* genome. For species whose *MAT* genes were not available from NCBI (Table A.2), various BLASTn and tBLASTn searches were conducted against their genomes using *MAT* genes from closely related species. For example, to search for the *MAT* locus of *N. discreta*, the three *MAT* genes from *N. crassa* were used in BLASTn searchers.

Table A.2: The accession numbers or Gene IDs for all genes used in this study.

Species	Gene	Accession Number/Gene ID
N. crassa	*matA-1 (MAT1-1-1)*	3880391
	matA-2 (MAT1-1-2)	3880488
	matA-3 (MAT1-1-3)	3880489
	*mfa-1 (*a-factor pheromone)	XM_011396530.1
	*ccg-4 (*α-factor pheromone)	XM_960001.3
	*pre1 (*a-factor receptor)	3872329
	*pre2 (*α-factor receptor)	3875871
N. discreta	*matA-1 (MAT1-1-1)* [a]	Appendix B
	matA-2 (MAT1-1-2) [a]	Appendix B
	matA-3 (MAT1-1-3) [a]	Appendix B

	mfa-1 (a-factor pheromone) [a]	Appendix B
	ccg-4 (α-factor pheromone) [a]	Appendix B
	pre1 (a-factor receptor) [a]	Appendix B
	pre2 (α-factor receptor) [a]	Appendix B
N. tetrasperma	*matA-1 (MAT1-1-1)*	XM_009857139.1
	matA-2 (MAT1-1-2)	XM_009857138.1
	matA-3 (MAT1-1-3)	XM_009856640.1
	mfa-1 (a-factor pheromone) [a]	Appendix B
	ccg-4 (α-factor pheromone)	XM_009858490.1
	pre1 (a-factor receptor)	AM749905.1
	pre2 (α-factor receptor)	AM904749.1
N. terricola	*matA-1 (MAT1-1-1)*	HE600070.1
	matA-2 (MAT1-1-2)	HE600070.1
	mata-1 (MAT1-2-1)	HE600070.1
	mfa-1 (a-factor pheromone) [a]	Appendix B
	ccg-4 (α-factor pheromone) [a]	Appendix B
	pre1 (a-factor receptor) [a]	Appendix B
	pre2 (α-factor receptor)	HE861764.1
N. pannonica	*matA-1 (MAT1-1-1)*	HE600067.1
	matA-2 (MAT1-1-2)	HE600067.1
	matA-3 (MAT1-1-3)	HE600067.1
	mata-1 (MAT1-2-1)	HE600067.1
	mfa-1 (a-factor pheromone) [a]	Appendix B
	ccg-4 (α-factor pheromone) [a]	Appendix B
	pre1 (a-factor receptor)	HE861745.1
	pre2 (α-factor receptor)	HE861760.1
N. africana	*matA-1*	HE600066.1
	matA-2	HE600066.1
	matA-3	HE600066.1
	mfa-1 (a-factor pheromone) [a]	Appendix B
	ccg-4 (α-factor pheromone) [a]	Appendix B
	pre1 (a-factor receptor)	HE861735.1
	pre2 (α-factor receptor)	HE861751.1
H. bhutanensis	*MAT1-2-1* [a]	Appendix B
	MAT1-2-7 [a]	Appendix B
	a-factor pheromone [a]	Appendix B
	α-factor pheromone [a]	Appendix B
	pre1 (a-factor receptor) [a]	Appendix B
	pre2 (α-factor receptor) [a]	Appendix B
H. omanensis	*MAT1-2-1*	KU950302.1
	MAT1-2-7	KU950303.1
	a-factor pheromone [a]	Appendix B
	α-factor pheromone [a]	Appendix B
	pre1 (a-factor receptor) [a]	Appendix B
	pre2 (α-factor receptor) [a]	Appendix B
H. moniliformis	*MAT1-2-1*	KU950299.1

	MAT1-2-7	KU950298.1
	a-factor pheromone [a]	Appendix B
	α-factor pheromone [a]	Appendix B
	pre1 (a-factor receptor) [a]	Appendix B
	pre2 (α-factor receptor) [a]	Appendix B
	MAT1-1-1	GAD92179.1 [b]
	MAT1-1-9 [a]	Appendix B
B. spectabilis	*ppgA* (α-factor pheromone) [a]	Appendix B
	preA (a-factor receptor) [a]	Appendix B
	preB (α-factor receptor) [a]	Appendix B
	MAT1-1-1 [a]	Appendix B
	MAT1-2-1 [a]	Appendix B
	MAT1-2-4 [a]	Appendix B
B. fulva	*ppgA* (α-factor pheromone) [a]	Appendix B
	preA (a-factor receptor) [a]	Appendix B
	preB (α-factor receptor) [a]	Appendix B
	MAT1-2-1 [a]	Appendix B
	MAT1-2-4 [a]	Appendix B
B. nivea	*ppgA* (α-factor pheromone) [a]	Appendix B
	preA (a-factor receptor) [a]	Appendix B
	preB (α-factor receptor) [a]	Appendix B
	MAT1-1-1 [a]	Appendix B
	MAT1-1-9 [a]	Appendix B
T. aurantiacus	*ppgA* (α-factor pheromone) [a]	Appendix B
	preA (a-factor receptor) [a]	Appendix B
	preB (α-factor receptor) [a]	Appendix B
	MAT1-2-1 [a]	Appendix B
	MAT1-2-4 [a]	Appendix B
T. crustaceus	*ppgA* (α-factor pheromone) [a]	Appendix B
	preA (a-factor receptor) [a]	Appendix B
	preB (α-factor receptor) [a]	Appendix B
	ppgA (α-factor pheromone)	7919755
Aspergillus flavus	*preA* (a-factor receptor)	7912518
	preB (α-factor receptor)	7914252

[a] This study
[b] Protein sequence

3.2 Pheromone genes

3.2.1 *α-factor pheromone*

Pheromone genes for which reference sequences existed were downloaded from

NCBI (Table A.2) and used in BLASTn queries against their respective genomes in order

6

to identify the gene's location as previously described for the *MAT* genes. For species where the pheromone genes were not available from NCBI (Table A.2), tBLASTn searches were conducted against their genomes using pheromone genes of closely related species. This method successfully identified all the α-factor pheromone genes in the *Neurospora* and *Huntiella* species.

Identification of the α-factor pheromone genes in the *Byssochlamys* and *Thermoascus* genomes was achieved using a microsynteny approach based on the location of the α-factor pheromone gene in the *A. flavus* genome. The proteins encoded by the genes directly flanking the *A. flavus* α-factor gene were used in local tBLASTn searches against the genomes of the *Byssochlamys* and *Thermoascus* species. The region between the identified flanking genes in these species was subjected to Fgenesh (Salamov and Solovyev, 2000) and manual annotation in order to identify the pheromone gene.

The predicted α-factor pheromone genes of all species considered in this study were translated and analyzed using the hydrophobicity and signal peptide predicting software programs, ExPASy ProtScale (Gasteiger et al., 2005) and Phobius (Krogh et al., 2007). Hydrophobicity was measured using the Kyte & Doolittle measure, with a window size of 9, a relative weight of window edges compared to window centers of 100%, a linear weight variation model and no scale normalization. Probability outputs from both programs were used to generate plots using the *ggplot2* package in R (Wickham, 2010; Wilkinson et al., 2005) (Figs S1 and S2). The potential mature α-factor repeats were manually annotated (Table A.3).

Table A.3: Structure and repeat sequence of the α-pheromones from each species.

[a] The gene sequence for the α-pheromone from *N. pannonica* was split across two contigs and had to be manually assembled. It is likely that sequence data is missing and thus there may be a peptide repeat missing from this schematic.

Species	Pheromone Structure	Repeat Sequence
N. crassa	---■-■-■-■-■---	■ QWCRIHGQSCW
N. discreta	---■-■-■-■-■---	■ QWCRIHGQSCW ■ QWCHIHGQSCW
N. tetrasperma	---■-■-■-■-■---	■ QWCRIHGQSCW
N. terricola	---■-■-■-■-■---	■ QWCRIHGQSCW ■ QWCQIHGQSCW
N. pannonica [a]	---■-■-■-■-■---	■ QWCRIHGQSCW ■ QWCQLHGQSCW
N. africana	---■-■-■-■-----	■ QWCRIHGQSCW ■ QWCRIRGQSCW
H. bhutanensis	---■-■-■-■-■-■-■-■-■---	■ DSNGGLPGELL ■ NSNGGLPGELL
H. omanensis	---■-■-■-■-■-■-■---	■ DSNGGLPGELL ■ NSNAGLPGELL ■ YSNAGLPGELL
H. moniliformis	---■-■-■-■-■-■---	■ DANGGLPGELF ■ DAWGGLPGELF
B. spectabilis	---■---	■ WCRRPGQPC
B. fulva	---■---	■ WCVDPGQVC
B. nivea	---■---	■ WCRRPGQPC
T. aurantiacus	------	N/A
T. crustaceus	---■---	■ WCSLNGQGC

3.2.2 *a-factor pheromone*

In order to identify the a-factor pheromone genes from the *Neurospora* and *Huntiella* genomes, we used the same method as was used for the α-factor pheromone gene

identification (Table A.2). These genes were all translated into their predicted proteins to confirm the presence of the conserved C-terminal CaaX domain (Table A.4)

Table A.4: Sequence of the a-pheromones from the *Neurospora* and *Huntiella* species. The C-terminal CaaX domains have been indicated in red.

Species	Pheromone Sequence
N. crassa	MPSTAASTKVPQTTMNFNGYCVVM
N. discreta	MPSTAASTKVPQTTMNFNGYCVVM
N. tetrasperma	MPSTAASTKVPQTTMNFNGYCVVM
N. terricola	MPSTAASTKVPQTTMNFNGYCVVM
N. pannonica	MPSTAASTKVPQTTMNFNGYCVVM
N. africana	MPSTAASTKVPQTTMNFNGYCVVM
H. bhutanensis	MAAIKNITSSKNAARGVDQSNPCNVMRGVDQSNPCAVMRGVDQSNPCTVM
H. omanensis	MAAIKNTTTSKNAARGVDQSNPCAVMRGVDQSNPCAVMRGVDQSNPCTVM RGVDQSNPCTLM
H. moniliformis	MPSIKNHTPSTKTSGNETIQPPTSNAGRGAIQSPINPITRGVTQAPPCNV MRGVTQAPPCNVM

A similar approach was not conducted for the *Byssochlamys* and *Thermoascus* genomes, because no a-factor pheromone has been identified in the genomes of any other Eurotiomycetes species and thus no pheromone sequences were available from closely related species. Thorough tBLASTn searches using all the a-factor pheromone sequences from (Dyer et al., 2003; Pöggeler, 2002) yielded no significant results. A method similar to the microsynteny-based approach used above also failed to produce usable results. Instead, all of the predicted coding sequences in the *Byssochlamys* and *Thermoascus* genomes were translated into putative protein products. These proteins were then filtered for those harboring the terminal CaaX

domain. These proteins were then screened to identify proteins with other similarities to previously identified a-factor pheromones.

3.3 Pheromone receptor genes

The receptor genes were identified using the same method as was used to identify the *MAT* genes (Table A.2). Thus, where sequences were available from NCBI, they were used to identify the genes in the genomes. Where sequences were not available, sequences of closely related species were used in BLASTn and tBLASTn searches as necessary. The a- and α-factor receptor genes from each of the 14 species were translated and subjected to functional domain discovery using the NCBI Conserved Domain Search (Marchler-bauer et al., 2015) to confirm their identity. These proteins were also subjected to hydrophobicity analysis using Phobius (Krogh et al., 2007) in order to identify putative transmembrane domains. Probability outputs from Phobius were used to generate plots using the *ggplot2* package in R (Wickham, 2010; Wilkinson et al., 2005).

4. Gene, protein and functional domain comparisons

All nucleotide and amino acid comparisons were conducted using the *Create Alignment* and *Create Pairwise Comparison* functions in CLC MainWorkbench V8.1 (CLC bio, Aarhus, Denmark). The gene, protein and functional domain alignments were conducted using the default settings, which includes gap open and gap extension costs of 50 and the "Very accurate (slow)" alignment setting. Pairwise comparisons were also conducted using the default settings and included gap, differences, distance, percent identity and identity comparisons.

5. RNA-seq mapping

Expression analysis was conducted using CLC Genomics Workbench V7.5 (CLC bio, Aarhus, Denmark). The raw data were filtered to retain only reads with a Phred score of at least 20 ($Q \geq 0.01$). The filtered reads were then mapped to the contigs containing the various genes of interest using the NGS Core Tool *Map Reads to Reference.* No masking was included, and the mapping settings were maintained at default. To ensure maximum mapping, the minimum length fraction and minimum similarity fraction values were set at 0.5 and 0.8, respectively. These relatively lenient mapping parameters ensured that reads spanning introns could successfully be mapped. This also ensured that successful mapping occurred despite the genome and transcriptomes originating from different isolates, as is the case for the majority of the species considered in this study. The resulting gene mappings were used to: 1) confirm or correct the gene models predicted by the various gene annotation programs and, 2) determine whether the various genes of interest are expressed in the two different sexual systems.

6. References

Dyer, P.S., Paoletti, M., Archer, D.B., 2003. Genomics reveals sexual secrets of *Aspergillus.* Microbiology 149, 2301–2303.

Ellison, C.E., Stajich, J.E., Jacobson, D.J., Natvig, D.O., Lapidus, A., Foster, B., Aerts, A., Riley, R., Lindquist, E.A., Grigoriev, I. V, Taylor, J.W., 2011. Massive changes in genome architecture accompany the transition to self-fertility in the filamentous fungus *Neurospora tetrasperma*. Genetics 189, 55–69.

Galagan, J.E., Calvo, S.E., Borkovich, K.A., Selker, E.U., Read, N.D., Jaffe, D., Fitzhugh, W., Ma, L., Smirnov, S., Purcell, S., Rehman, B., Elkins, T., Engels, R., Wang, S., Nielsen, C.B., Butler,

J., Endrizzi, M., Qui, D., Ianakiev, P., Bell-pedersen, D., Nelson, M.A., Werner-washburne, M., Selitrennikoff, C.P., Kinsey, J.A., Braun, E.L., Zelter, A., Schulte, U., 2003. The genome sequence of the filamentous fungus *Neurospora crassa*. Nature 422, 859–868.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy Server, in: Walker, J. (Ed.), The Proteomics Protocols Handbook. Springer: Humana Press Inc., Totowa, NJ, pp. 571–608.

Gioti, A., Stajich, J.E., Johannesson, H., 2013. *Neurospora* and the dead-end hypothesis: Genomic consequences of selfing in the model genus. Evolution (N. Y). 67, 3600–3616.

Krogh, A., Sonnhammer, E.L.L., Ka, L., 2007. Advantages of combined transmembrane topology and signal peptide prediction - The Phobius web server. Nucleic Acids Res. 35, W429–W432.

Marchler-bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Bryant, H., 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res. 43, D222-226.

Mcclendon, S.D., Batth, T., Petzold, C.J., Adams, P.D., Simmons, B.A., Singer, S.W., 2012. *Thermoascus aurantiacus* is a promising source of enzymes for biomass deconstruction under thermophilic conditions. Biotechnol. Biofuels 5, 1–9.

Pöggeler, S., 2002. Genomic evidence for mating abilities in the asexual pathogen *Aspergillus fumigatus*. Curr. Genet. 42, 153–160.

Prefecture, K., 2014. Draft genome sequence of the formaldehyde-resistant fungus *Byssochlamys spectabilis* No. 5 (Anamorph *Paecilomyces variotii* No. 5) (NBRC109023). Microbiol. Resour. Announc. 2, e001162-13.

Salamov, A.A., Solovyev, V. V, 2000. *Ab initio* gene finding in *Drosophila* genomic DNA. Genome Res. 10, 516–522.

Schuerg, T., Prahl, J.P., Gabriel, R., Harth, S., Tachea, F., Chen, C.S., Miller, M., Masson, F., He, Q., Brown, S., Mirshiaghi, M., Liang, L., Tom, L.M., Tanjore, D., Sun, N., Pray, T.R., Singer, S.W., 2017. Xylose induces cellulase production in *Thermoascus aurantiacus*. Biotechnol. Biofuels 10, 1–11.

Singer, S.W., n.d. Genome assembly of *Thermoascus aurantiacus*. Joint Genome Institute.

Stanke, M., Waack, S., 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19, 215–225.

van der Nest, M.A., Beirn, L.A., Crouch, J.A., Demers, J.E., de Beer, Z.W., De Vos, L., Gordon, T.R., Moncalvo, J.-M., Naidoo, K., Sanchez-Ramirez, S., Roodt, D., Santana, Q.C., Slinski, S.L., Stata, M., Taerum, S.J., Wilken, P.M., Wilson, A.M., Wingfield, M.J., Wingfield, B.D., 2014. Draft genomes of *Amanita jacksonii*, *Ceratocystis albifundus*, *Fusarium circinatum*, *Huntiella omanensis*, *Leptographium procerum*, *Rutstroemia sydowiana*, and *Sclerotinia echinophila*. IMA Fungus 5, 472–485.

Van Der Nest, M.A., Bihon, W., De Vos, L., Naidoo, K., Roodt, D., Rubagotti, E., Slippers, B., Steenkamp, E.T., Markus Wilken, P., Wilson, A., Wingfield, M.J., Wingfield, B.D., 2014. Draft genome sequences of *Diplodia sapinea*, *Ceratocystis manginecans*, and *Ceratocystis moniliformis*. IMA Fungus 5, 135–140.

Wang, Z., Lopez-giraldez, F., Lehr, N., Common, R., Trail, F., Townsend, J.P., 2014. Global gene expression and focused knockout analysis reveals genes associated with fungal fruiting body development in *Neurospora crassa*. Eukaryot. Cell 13, 154–169.

Wickham, H., 2010. A Layered grammar of graphics. J. Comput. Graph. Stat. 19, 3–28.

Wilkinson, L., Anand, A., Grossman, R., 2005. Graph-Theoretic Scagnostics. Proc. EEE Symp.

Inf. Vis. 1, 21.

Wilson, A.M., van der Nest, M.A., Wilken, P.M., Wingfield, M.J., Wingfield, B.D., 2018. Pheromone expression reveals putative mechanism of unisexuality in a saprobic ascomycete fungus. PLoS One 13, e0192517.

Wingfield, B.D., Duong, T.A., Hammerbacher, A., van der Nest, M.A., Wilson, A., Chang, R., Wilhelm de Beer, Z., Steenkamp, E.T., Markus Wilken, P., Naidoo, K., Wingfield, M.J., 2016. Draft genome sequences for *Ceratocystis fagacearum*, *C. harringtonii*, *Grosmannia penicillata*, and *Huntiella bhutanensis*. IMA Fungus 7, 317–323.

Yin, G., Hua, S.S.T., Pennerman, K.K., Yu, J., Bu, L., Sayre, R.T., Bennett, J.W., 2018. Genome sequence and comparative analyses of atoxigenic *Aspergillus flavus* WRRL 1519. Mycologia 110, 482–493.

SUPPLEMENTARY FILE B: SEQUENCES OF INTEREST

Note: Because these sequences were extracted and predicted from publicly available genome and transcriptome data, they could not be resubmitted to a publicly available database.

N. discreta, *matA-1*, MAT1-1-1 protein

Gene sequence:

```
ATGTCGGGCGTCGACCAAATTGTCAAGACATTTGCTGACCTCGCTGAGGACGACCGCGAAGCGGCA
ATGAGAGCTTTCTCAACGATGATGCGTACCGAACCTGTTCGCCAAATCCCCGCGACAAAGAAGAAG
GTCAACGGCTTCATGGGTTTCAGATGTAAGTTAAATCTGAATCTTGTCGATAATCCATGCTGACTG
CTCTTCATTTCAGCGTACTATTCCCCGCTCTTCTCTCAGCTCCCGCAAAAGGAGAGATCGCCCTTT
ATGACCATTCTCTGGCAGCACGATCCCTTCCACAACGAATGGGATTTCATGTGCTCGGTGTATTCG
TCAATCCGTACCTATCTTGAGCAGGAGAAGGTTACTCTGCAACTTTGGATTCACTATGCTGTTGGC
CATCTGGGAGTGATTATCCGCGACAACTACATGGCATCGTTTGGCTGGAACCTCGTCCGTCTGCCC
AACGGCACACACGACCTCGAGCGCACCGCTCTTCCCTTGATTAGGCACAATCTCCAGCCCATGAAC
GGCCTATGCCTGCTCACAAAGTGCCTCGAGAGCGGATTGCCTCTTGCCAATCCTCACTCTGTCATC
GCCAAGCTTTCAGATCCTAGCTACGACATGATCTGGTTCAACAAGCGTCCTTACAGTCAGCAGAAA
CATTTTGGTCAAACTGACGATTCTGAACTCGAAGTGTCGGCGATGTTCCCTCGCAATTACGCAGTC
GCCGCAGAGGCAGATGGTATCACCAATCTTCCTCTCTCCCATTGGATTCAGCAGGGAGATTTCGGT
ACTGAGTCCGGATACTCAGCTCATTTTGAGACTTTGTTGGATTCTATTCTCGAGAATGGACACGCC
TCCAGCAATGACCCCTACAACATGGCTCTGGCTATGGATGTTCCTATGATGGGTTAG
```

CDS annotation:

```
join(1..157,212..915)
```

Protein:

```
MSGVDQIVKTFADLAEDDREAAMRAFSTMMRTEPVRQIPATKKKVNGFMGFRSYYSPLFSQLPQKE
RSPFMTILWQHDPFHNEWDFMCSVYSSIRTYLEQEKVTLQLWIHYAVGHLGVIIRDNYMASFGWNL
VRLPNGTHDLERTALPLIRHNLQPMNGLCLLTKCLESGLPLANPHSVIAKLSDPSYDMIWFNKRPY
```

SQQKHFGQTDDSELEVSAMFPRNYAVAAEADGITNLPLSHWIQQGDFGTESGYSAHFETLLDSILE

NGHASSNDPYNMALAMDVPMMG*

N. discreta, *matA-2*, MAT1-1-2 protein

Gene sequence:

ATGGATCTCATCAACATGCAACCTAGAAAGTCAGAGCAACCGGTTAGGTTTGAAGAAAACCGTGTC

TCTAGCCAGGAAGGCCAGGATCTCGAAGTGATGTACAAGGTAACTATTTTCTGACTCAAAAAAAAA

CTCATTTACTTGTCACTGATGAATTGGTCAGAAACTCCATCAGCTACAGGCTAGGCTTTCCCGTTC

AGTTCTTTCAGAGGCAATCAAGGAGTTCGAAGAGAATCTTCGGTGTCTCTTCTATGAAGCCAAGCT

CTTACTATGCACAACACGAACGAAGTATCGCCAAAGCTGGTTCGGTTCTAGCAACGAGTTTGGGTC

TATCGACGAGAGCAGGATTATCAAGGCATCGTGCTGTATCATTGAGTCGGCCAACACAATTCTCAA

TTTTCTCTCATTTATTGAGAAGAAACGAGGATTGCCATCAGGCGGAGATCAAAGACTCCAACAAGC

TGCCTACAAAGGCCAGCAGTTTGCGTTTCGCCTCCTTCGCTCACTTACATGCCACAAAGATGCTCA

GGAGGCTCCCGGAAAGGAATTTGGCTTGGCCTACGGAAAAGAGGTGTATGTAATGGATGGCCATCT

TTTGCACAGGTCAAAGCAAGAGGTCGTGGGGCAGGCGGGAGGAAGAAACTGGCATATCGACCATAC

TCTCCACCCTTTGAGGCGCGTCCCAGGCACCCCATGGCACAAGTTCTTTGGCAATCTCGAAGTCGA

CGCCGACAAGCAACTTCGCCTCTTCGATGATGGTACGTCTGTCGACAATGATCGAGTCGGACCTCG

GAAGTTCTTTGTAGTTGTTCCGGAAACTGCTGAATTCATTTTGGACGAAATCAGGAGCGAGCATCA

GAGAGTCGCTACGATTCACACAGAGGTGATTACTTCAACGCGTCGAAACAGACCAAAACTCGCACG

ACTGACTTATGGTAGAGTGGACATGCCCAGCCGCCAGTACCGACATCTATTCAGCAAGAGGTAAGT

TCCTTTGGCCCGTCTCAACGTCATGCTATCCCATCGTGCAGAGCACAATGCACGCTCTTTCGACTC

ACTAATGATGACTGACATTATTGAAGGCTCTCCTCAGGAAGTTGGACTTTGCCATGACAACACCGT

TCCATGGTTACGTTGTAGAAGGACAACCTGAGATTGTGTTCCATCATGAAGGCTTGCGCCAGGTTC

GTATGATCCTGCCCACTGCTCACGGATGATAATGTGCTAACAACTGATCAACAGGTCCCCATTGAC

TACAGCCAGGAGCGCCAACTCAGCATCCTCTCCCATGTTTTCACTCGACCCGCACTTTGGGGGGAA

GGTCTGGAGCTTGCGGATAACTTTGACCCGCGAGACGGTGTGCAGCAAGAAGAGCACATCTATTAC

ACCTGA

CDS annotation:

```
join(1..105,164..883,940..984,1083..1184,1243..1392)
```

Protein:

```
MDLINMQPRKSEQPVRFEENRVSSQEGQDLEVMYKKLHQLQARLSRSVLSEAIKEFEENLRCLFYE

AKLLLCTTRTKYRQSWFGSSNEFGSIDESRIIKASCCIIESANTILNFLSFIEKKRGLPSGGDQRL

QQAAYKGQQFAFRLLRSLTCHKDAQEAPGKEFGLAYGKEVYVMDGHLLHRSKQEVVGQAGGRNWHI

DHTLHPLRRVPGTPWHKFFGNLEVDADKQLRLFDDGTSVDNDRVGPRKFFVVVPETAEFILDEIRS

EHQRVATIHTESGHAQPPVPTSIQQEALLRKLDFAMTTPFHGYVVEGQPEIVFHHEGLRQVPIDYS

QERQLSILSHVFTRPALWGEGLELADNFDPRDGVQQEEHIYYT*
```

N. discreta, *matA-3*, MAT1-1-3 protein

Gene sequence:

```
ATGTCTGCCTTAGACGTTGATGCAATCAGCGACATCGCACCCGGTCTCAGCCCTGTAACTGCCCTT

CACTATGGCAGGATCCAGGTAATGCTGTTTAGATCCCATCTGGCCGATTTTGCTGAAGAGGATCTT

GTCTATGCGATGGACAACTCAGCGTGAGTCTTGCCCCAAGGAAAGCTTCTTTCAGATCTCATAGCT

GACTCGTTGTAGTGTCGTATTTGGCGAAGAGGCTTTGCTTATGGTTGCCCCCGATGAGACAAGCAT

CGCGATTTGCACGTATCCAGTTGGACTCATGATGGTGGAATGGGGAAACTGGGACATACTCGCGGT

TTCGCCCCCTCGTAAGTTCTCCAACATTGGGCGCCCTGCACTCAAGCTAATGAGACATTCCAGCGC

AAACTCCTACCATTCCCGGCGAGAACACCTTGAGCATCTCAAATCAAGGAGGTGCTAATCCCGAGC

AGCAAGAGCAAAGCTCCCATACCGTCGACATGACTCTGCCCATCAACTTTTTTGAACAGAGTTCGG

TAACCCAGAGCAATGGTACTAGCCGCCCCCGCAACCAGTTCGTTCTATATTACCAATGGCTCTTGG

ATACTCTGTTCTCCGAAGATCCAAGTCTATCAGCTCGCGATATTTGTATGTGAACGAACTCTCCCG

TCCTGTATGGCATTTATTGACAATATCTACACAGCTCAAGTTGTTGCGGGCTTGTGGAACAACGAG

CACCCCGCGGCAAAAGCTCGCTTCATGGAACTGGCGGAAATGGAGGTTCAGCGTCACCGAGCTGAG

AATCCTCACCTTTACCCCGACCAGTCGCGATTTCCCACGACTGACCCCGTTCCTCCTCGCATGAGA

TATCCTTGCGTAATCTCGCCTGAAGATCGGCAACGAATTCTGCGGATGCTCGACTTTGTCTGGGAA
```

```
GAGTCGAACGGTCAGCTGGCTGCCGAAGAAGCCGCATTTAACGATACCGAACAAACCCAAGAAGCC

GAAGAGGTCGACCCTTTCCCCGACTACGAGTGGGAGGAGCCCAACCAGATTATTGACATGTCGACC

GACCTGTCCGTAGCGCAGGACCCGGACTTCATGATGACGGAGGACGACTCCATGGGATTCCTTCTC

AAGCAGGCGTGCTGA
```

CDS annotation:

```
join(1..155,211..341,394..639,695..1137)
```

Protein:

```
MSALDVDAISDIAPGLSPVTALHYGRIQVMLFRSHLADFAEEDLVYAMDNSAVVFGEEALLMVAPD

ETSIAICTYPVGLMMVEWGNWDILAVSPPPQTPTIPGENTLSISNQGGANPEQQEQSSHTVDMTLP

INFFEQSSVTQSNGTSRPRNQFVLYYQWLLDTLFSEDPSLSARDISQVVAGLWNNEHPAAKARFME

LAEMEVQRHRAENPHLYPDQSRFPTTDPVPPRMRYPCVISPEDRQRILRMLDFVWEESNGQLAAEE

AAFNDTEQTQEAEEVDPFPDYEWEEPNQIIDMSTDLSVAQDPDFMMTEDDSMGFLLKQAC*
```


N. discreta, mfa-1, a-factor pheromone

Gene sequence:

```
ATGCCTTCCACCGCTGCTTCCACCAAGGTCCCCCAGACCACCATGAACTTCAACGGTTACTGCGTT

GTTATGTAA
```

CDS annotation:

```
(1..75)
```

Protein:

```
MPSTAASTKVPQTTMNFNGYCVVM*
```


N. discreta, ccg-4, α-factor pheromone

Gene sequence:

```
ATGAAGTTCACTCTCCCTCTTGTCATCTTCGCCGCCGTGGCCTCCGCCACCCCGGTCGCCCAGCCA

AACGCTGAGGCCGAAGCCCAGTGGTGCCGAATCCACGGCCAGTCCTGCTGGAAGGTCAAGCGTGTT
```

```
GCCGATGCCTTCGCCAACGCCATCCAGGGCATGGGTGGTCTCCCGCCTCGCGATGAGTCCGGCCAC

CAGCCCGCTCAGGTCGCCAAGCGCCAGGTTGACGAGCTTGCCGGCATCATTGCCCTCACTCAGGAG

GACGTCAACGCCTACTACGACTCGCTTGGCCTCCAGGAAAAGTTCGCCCCATCCACCGAGGAGGAG

AAGAAGACCGAGAAGGTCGCCAAGCGTGAAGCCGAGGCCGAGGCACAATGGTGCCACATCCACGGC

CAATCGTGCTGGAAGAAGCGTGAGGCCGAAGCCCAGTGGTGCCGCATTCACGGTCAGTCCTGCTGG

AAACGTGACGCCCTCCCCGAGGCCGAGCCCCAGTGGTGCCGCATTCACGGCCAGTCCTGCTGGAAG

AAGCGTGACGCCGCTCCCGAGGCTGCTCCCGAGGCTGCTCCCGAGGCTGAGGCCAACCCGCAATGG

TGCCGCATCCACGGCCAGTCCTGCTGGAAGGCCAAGCGCGCCGCCGAGGCCGTCATGACCGCCATC

CAGTCCGCCGAGGCCGAGTCTGCCCTCCTTCTCCGTGACAATACCTTCAGCCCCGTCGACCGGGTC

GGCAAGCGCGATACTTGCTGGAACGTCAGGCTCCCCGGTGGTGTCAAGAGCTCCTGCTGGAAGCGT

GATGCCTCCCCAGAGGCGGCTTGCAATGCTCCCGACGGCGCTTGCACCAAGGCCACCCGTGACTTG

CACGCCATGTACAACGTCGCTCGTGCCATCCTCACTGCTCACTCCGATGAGAACTAG
```

CDS annotation:

```
(1..915)
```

Protein:

```
MKFTLPLVIFAAVASATPVAQPNAEAEAQWCRIHGQSCWKVKRVADAFANAIQGMGGLPPRDESGH

QPAQVAKRQVDELAGIIALTQEDVNAYYDSLGLQEKFAPSTEEEKKTEKVAKREAEAEAQWCHIHG

QSCWKKREAEAQWCRIHGQSCWKRDALPEAEPQWCRIHGQSCWKKRDAAPEAAPEAAPEAEANPQW

CRIHGQSCWKAKRAAEAVMTAIQSAEAESALLLRDNTFSPVDRVGKRDTCWNVRLPGGVKSSCWKR

DASPEAACNAPDGACTKATRDLHAMYNVARAILTAHSDEN*
```

N. discreta, *pre1*, a-factor pheromone receptor

Gene sequence:

```
ATGAACAACACTGATTCCTGGGCAAGTGTCGCCAACATCACCACCTCCTACACCACCACACCTGAA

GAACGACTTGGCCCTCCCGCTCCTTACACCAATATTGGTCTTCAGGTCAACCTCTTCTTCAGAGTC

TTTCTGGGAATTCTCGGAATACTCATTCCTCTTGTGCCTGCAAGGCTGTTATGGATCAATGGCGAA

TTTTCTGCCACGGTACATTGCATGTCAACAGTCACGCTCAACTTCTTTTACGTCGTCAACTCGCTC
```

ATATGGCGAGACAACAATGTCAAGAAATGGTATGCCGGCTATGGGTGGTGCGATTTTCATACTTAC

GTCTTCTTTGCGGTTGAAACTATCTTCCACACCACTCTCTTCGACATCATGCTCGGTCTTGCCATC

AAGATCGGTAGCCCGCGAGTCACCAGCCTCAGTCCTAAGGAAAAGAAACGCAAGGACCGAGTTTCG

GCCCTCATCATCTTCGCCAATCCGGTCTTACAGGTCTTGCTCACCTACTTTCTCATTACGCAACGG

TACGACGTCTTCACTCTTGCCGGTTGCAATGCCATCTTCGATCCCAATGGCGTCTTCTTCGTTTTC

TTCATTCTCCCATCGCCAGTCTTCACGCTTGGGGCGGCTGGCTTGGCAGGTAAGATACAAAACCAA

GTCCATGGCATGTTTCGCAATCGCTCACAATGTACACAGGGGTTTGTTTTTACAAGTACCGGCAGC

TGGAGAAACTCACCCGCGAAGTCTTACCAAGTGATGACAGTATACGAACCGCGAGACAAAAACGCC

TGAGGAGAAAGCTCTACTTTCTCACTCTCTCTATTCTCGTCTTAGTCGTGCCGACTGTTTGCGTCT

TCTTCGTCTTCAACCTCATTCAAGGTTGGCCGTGGTCACTGCCGTTCGACCTCCATCGCATTCACG

CCAACATCAACTTTGTCTCGTTCACAACGACGGAAAGAATGAAAGTCTCCGCAGTCCTTACCAACT

ATGTGCCAGTGGTGAGCTCAGCGGCAATTTGCATCACCTTCGGCACAACAGTCGAGGCATATAACC

AATACCGCTTAGTTCTCGTCTTCCTTGGCTTAGGAAAAGTTTGGCCCAAGCTGTATCAGGAATATG

ACCCCGATGAGTCTGGACCACCATCTGAGCTTAGCACGCAAACAAGTTCGAAATCGTGGTGGTCAT

CCATAACGACAAACAACAAAAGGGGTACAGTCACAAGGTATAAGATCACACACCTCACAACATAAC

TATCACATTGGATCATCTGGATGCTGACGTACAGTTGAATAAGCAGCCAGTGTGGAAATGACTTCT

CTATCCTCCCTATTACTGAGGACATACCACTCACAAACCAATCTAAACCTGCACAAACGGGTGAAG

GCTCCTATACCGTCCCGGGTTATGCATACGATGAACTCGGAAAAGAAATCTCGAACCAGTATTCTA

CCCAGTCCTCAGCCACAGCAAAATCCGACCATAATCCATGGCCCGATCTGTCAGACAATCACCCCC

CTCCAGCACGCAACCCATGGTATTTCCGACCGAATGCTTTCCACGTCCCGATCACATTGCCAATGC

CACTGTCGCCCATTTACATACCGTCTCTGCATGCGGGTCAAGCAAAGCAGACGGAGGGAACGCAGA

AACGGACGACAGTCAACAGGCTTTCTCAAGGCTACACCGCACTCAGTAGCTCCTCCACAACAATCC

TACCCCCCTCGAATTCACCTACCTGCCCTCTTCCCCCGATTCCAATCTCCACATCCTCCTCAAGAC

ATTCACCACAGGAGGAACGACACTGGGCGACGAGAAAGACACAAAAACAGAAAGAACCAGACACGC

AAAATAGCACAAGCAATGGGACAACCAAGAACCCATTGACACCAGCGGATCCATGGAGTCCCTTTG

CTTCCAGTCACACCAACACTACCCTGGGAGTTGACACACGCGTGTGGTCCCAGAAGGATAACGACG

ACTCTGCGGGTCCTACCCAAGACTCACCGATCTCCCATGCTTCTCGTTCAATCACGGGACCCCAGC

TCTCACCGCCGCCATCATACAGTGTCGCTGTTGGTCAAGAATCCTCCTTTGCTTCCGCTGCCGATC

```
AATCGGCAGGATTCGAAATCAGTCGGCCTGAAAATCAACAGATTGGGACGGCGTTAGGACATCCAC

CGAGAGGCGGGGTGGTTAGAGTAGAGACCCATATAGCGAGTGAGATTGAGGTCATTAGCAGTCTTG

GGCGTGATGATGGTAATGATCGTGTTGGTGATCAACGAGGTGGAAGGGGGTGGTAG
```

CDS annotation:

```
join(1..643,700..1225,1301..2300)
```

Protein:

```
MNNTDSWASVANITTSYTTTPEERLGPPAPYTNIGLQVNLFFRVFLGILGILIPLVPARLLWINGE

FSATVHCMSTVTLNFFYVVNSLIWRDNNVKKWYAGYGWCDFHTYVFFAVETIFHTTLFDIMLGLAI

KIGSPRVTSLSPKEKKRKDRVSALIIFANPVLQVLLTYFLITQRYDVFTLAGCNAIFDPNGVFFVF

FILPSPVFTLGAAGLAGVCFYKYRQLEKLTREVLPSDDSIRTARQKRLRRKLYFLTLSILVLVVPT

VCVFFVFNLIQGWPWSLPFDLHRIHANINFVSFTTTERMKVSAVLTNYVPVVSSAAICITFGTTVE

AYNQYRLVLVFLGLGKVWPKLYQEYDPDESGPPSELSTQTSSKSWWSSITTNNKRGTVTSQCGNDF

SILPITEDIPLTNQSKPAQTGEGSYTVPGYAYDELGKEISNQYSTQSSATAKSDHNPWPDLSDNHP

PPARNPWYFRPNAFHVPITLPMPLSPIYIPSLHAGQAKQTEGTQKRTTVNRLSQGYTALSSSSTTI

LPPSNSPTCPLPPIPISTSSSRHSPQEERHWATRKTQKQKEPDTQNSTSNGTTKNPLTPADPWSPF

ASSHTNTTLGVDTRVWSQKDNDDSAGPTQDSPISHASRSITGPQLSPPPSYSVAVGQESSFASAAD

QSAGFEISRPENQQIGTALGHPPRGGVVRVETHIASEIEVISSLGRDDGNDRVGDQRGGRGW*
```

N. discreta, pre2, α-factor pheromone receptor

Gene sequence:

```
ATGGCCTCCTCCGTCGACATCTACAGCGGCATAACCCAATCCCTCAACTCCACCCACGCCACTCTC

ACCCTCCCCATCCCCCCCGCCGACCGCGACCACCTCACGAACCAAGTCCTCTTCTTCTTCGACAAC

TACGGCCACCTCCTCAACGTAACCACCACCGAAGTCGACACCCTCACCGACATGGTCATCTCCACC

GCCATCAACTACGCCACCCAAATCGGCGCCACCTTCATCATGCTCGCCATCATGCTTCTCATGACG

CCCCGCCGCCGTTTCAAGCGTCTACCCACCATCATCAGCGTCTTGGCCCTTTGCATGAACCTGATC

CGCGTCGTCCTGCTTGCGCTCTTCTGCCCGTCGCACTGGACCGATTTCTACGTCTTGTACTCGGGC

GACTGGCATTCCGTGCCCCGCAGCGACATGCAAATCAGCGTGGCAGCCACGGTGCTCAGCATCCCC
```

```
GTCACGGCCCTGCTGCTGTCGGCGCTCATGGTGCAAGCGTGGAGTATGATGCAGCTGTGGACACCG

CTATGGCGGGCCCTGGTCGTGTTGGTCTCTGGTCTGCTTTCGTTCGCGACCGTGGCGCTTTCGTTT

GCCAATTGCATTTTTCAAGCCAAGAATATCCTGTATGCGGACCCGTTCCCGCCGTATTGGGTGCGC

AAGCTGTACCTGGCGATGACGACGGGGTCGATTTCGTGGTTTACCTTCTTGTTCATGATTCGCTTG

GTCATGCATATGTGGACTAACAGGAGTATCTTGCCGAGTATGAAGGGGTTGAAAGCCATGGATGTG

TTGATTATTACTAACTCCATCTTGATGCTGATTCCGGTCTTGTTTGCCGGGCTGGAGTTCCTGGAC

TCGGCCAGTGGGTTTGAGAGCGGTTCGCTTACGCAGACGAGCGTCGTCATTGTCTTGCCGCTTGGC

ACGCTGGTTGCGCAACGGATTGCTACGAGGGGATACATGCCTGACTCTTTGGAGGCTTCTGGACCT

AACGGAAGTTTGCCGCTGAGCAATTTGAGTCTTACCGGTGGCGGTGGTGGTGGTGGAGGTGGTGGT

CACCAAGATAAAGGAAATGGAGGAGGAGGAGGAGGTAGACAAAATGGCACCATCCCTCCCACCACC

ACCAATAACAACAACAACATAGCAACCACCTTTTCTTCCTCCATCGCGTGCAGCGGCATCAGCTGC

CTGCCCAAGGCAAAGCGCATGACGGCCAGCTCGGCATCTTCCAGTCAACGACCCCTTTTGACAACG

ACCAACTCGACCATAGCCAGCAATGACTCGTCGGGTTTCCCGAGTCCAAGCGTCCACAATGCCACG

CAATACCAATATGCGATGAACATGAACATGAACATGCCCCACTACCACCCACCTGTTCCTTATCCT

GGGTACAACAGCCGTAACACGGGCGTTACCTCTCACATTGCCAGCGACGGCGGACACCACCAGGGT

ATGAACCGACATCCGAGCGTCGACCACTTTGATCGCGAGCTGGCGAGGATAGATGATGAAGATGAT

GTGGACAACCAAGATGTGTACCCCTTTGCGTCATTCGAGAAGGCGGTCATGGATGGCGATGATAAT

GATGATGATGTAGAGAGGGGGAGGGGGAGAGCCCTACCTCCGCCGCCGCTTGGGGGAGTGAGAGTG

GAGAGAAGGATTGAGACTAGGAGTGAGGAGAGGATGCCGTCGCCTAATCCGTTGGGAGTCACAAAG

CCGAGGTCTTTTGAGTAG
```

CDS annotation:

```
(1..1734)
```

Protein:

```
MASSVDIYSGITQSLNSTHATLTLPIPPADRDHLTNQVLFFFDNYGHLLNVTTTEVDTLTDMVIST

AINYATQIGATFIMLAIMLLMTPRRRFKRLPTIISVLALCMNLIRVVLLALFCPSHWTDFYVLYSG

DWHSVPRSDMQISVAATVLSIPVTALLLSALMVQAWSMMQLWTPLWRALVVLVSGLLSFATVALSF

ANCIFQAKNILYADPFPPYWVRKLYLAMTTGSISWFTFLFMIRLVMHMWTNRSILPSMKGLKAMDV

LIITNSILMLIPVLFAGLEFLDSASGFESGSLTQTSVVIVLPLGTLVAQRIATRGYMPDSLEASGP
```

```
NGSLPLSNLSLTGGGGGGGGGGHQDKGNGGGGGGGRQNGTIPPTTTNNNNNIATTFSSSIACSGISC

LPKAKRMTASSASSSQRPLLTTTNSTIASNDSSGFPSPSVHNATQYQYAMNMNMNMPHYHPPVPYP

GYNSRNTGVTSHIASDGGHHQGMNRHPSVDHFDRELARIDDEDDVDNQDVYPFASFEKAVMDGDDN

DDDVERGRGRALPPPPLGGVRVERRIETRSEERMPSPNPLGVTKPRSFE*
```

N. tetrasperma, mfa-1, a-factor pheromone

Gene sequence:

```
ATGCCTTCCACCGCTGCTTCCACCAAGGTCCCCCAGACCACCATGAACTTCAACGGCTACTGCGTT

GTTATGTAA
```

CDS annotation:

```
(1..75)
```

Protein:

```
MPSTAASTKVPQTTMNFNGYCVVM*
```

N. terricola, mfa-1, a-factor pheromone

Gene sequence:

```
ATGCCTTCCACCGCTGCTTCCACCAAGGTCCCCCAGACCACCATGAACTTCAACGGCTACTGCGTT

GTTATGTAA
```

CDS annotation:

```
(1..75)
```

Protein:

```
MPSTAASTKVPQTTMNFNGYCVVM*
```

N. terricola, ccg-4, α-factor pheromone

Gene sequence:

```
ATGAAGTTCACCCTCCCTCTTGTCATCTTCGCCGCCGTGGCCTCCGCCACCCCGGTCGCCCAGCCT
AACGCTGAGGCCGAAGCCCAGTGGTGCCAGATCCACGGCCAGTCCTGCTGGAAGGTCAAGCGTGTT
GCCGATGCCTTCGCCAACGCCATCCAGGGTATGGGTGGTCTCCCGCCCCGCGACGAGTCCGGCCAC
CAGCCCGCCCAGGTCGCCAAGCGCCAGGTCGACGAGCTCGCCGGCATCATCGCCCTCACCCAGGAG
GACGTCAACGCCTACTACGACTCCCTCGGCCTGCACGAGAAGTTCGCTCCCTCCACCGAGGAGGAG
AAGAAGACCGAGAAGGTCGCCAAGCGCGAGGCCGAGGCCGAGGCGCAATGGTGCCGCATCCACGGC
CAGTCGTGCTGGAAGAAGCGTGAGGCCGAAGCTCAGTGGTGCCGCATCCACGGCCAGTCCTGCTGG
AAGCGTGACGCTCTCCCCGAGGCCGAGCCCCAGTGGTGCCGCATCCACGGTCAATCTTGCTGGAAG
AAGCGCGATGCTGCTCCCGAGGCTGCTCCCGAGGCCGAGGCCAACCCGCAATGGTGCCGCATCCAC
GGCCAGTCCTGCTGGAAGGCCAAGCGCGCCGCCGAGGCCGTCATGACCGCCATCCAGTCCGCCGAG
GCCGAATCTGCTCTTCTCCTCCGTGACACCACCTTTAGCCCCGTCGACCGCGTCGGCAAGCGCGAT
GCCCAGTGGTGCAATGTCAAGTTCCCCGGCGGTGTTAAGGGCAACTGCTGGAAGCGTGATGCCTCC
CCCGAGGCGGCTTGCAACGCTCCCGATGGTGCTTGCACCAAGGCCACCCGTGACTTGCACGCCATG
TACAACGTCGCTCGTGCCATCGTCACTGCTCACTCCGATGAGAACTAG
```

CDS annotation:

```
(1..906)
```

Protein:

```
MKFTLPLVIFAAVASATPVAQPNAEAEAQWCQIHGQSCWKVKRVADAFANAIQGMGGLPPRDESGH
QPAQVAKRQVDELAGIIALTQEDVNAYYDSLGLHEKFAPSTEEEKKTEKVAKREAEAEAQWCRIHG
QSCWKKREAEAQWCRIHGQSCWKRDALPEAEPQWCRIHGQSCWKKRDAAPEAAPEAEANPQWCRIH
GQSCWKAKRAAEAVMTAIQSAEAESALLLRDTTFSPVDRVGKRDAQWCNVKFPGGVKGNCWKRDAS
PEAACNAPDGACTKATRDLHAMYNVARAIVTAHSDEN*
```

N. terricola, *pre1*, a-factor pheromone receptor

Gene sequence:

```
ATGAACAACACAGATTCCTGGGCAAGTGTCGCCAACAGTACCACCTCCTACACCACCACACCCGAA
GAACGACTTGGCCCTCCCGCTCCTTACACCGATATTGGTCTCCAAGTCAACCTCTTCTTCAGAGTC
```

TTTCTAGGAATCCTCGGAATTCTTATTCCTCTCGTTCCTGCGAAGCTATTATGGATCAACGGCGAA

TTTGGGGCCACGGTACATTGCATGTCAACAGTCACACTCAACTTCTTTTACGTCGTCAACTCGCTC

ATATGGCGAGACAACAATGTCAAGAAGTGGTATGCTGGCTATGGGTGGTGCGATTTCCATACTTAC

GTCTTCTTCGCGGTTGAAACCATCTTTCACACCACTCTCTTCGACATCATGCTCGGTCTTGCCAAC

AAGATCGGTAACCCGCGAGTCACCAGCCTCAGTCCCAAGGAAAAGAAACGCAAAGACCGAATCTCG

GCCCTGATCATCTTCGGCAATCCGGTCCTACAGGTCTTGCTCACTTACTTTGTCATTACTCAACGG

TACGACATCTCAACCCTTGCCGGTTGCAATGCCATCTTCGACCCCAATGGCGTCTTCTTCGTTTTC

TTCATTCTCCCATCGCCAGTCTTCACGCTTGGTGCGGCAGGCTTGGCAGGTAAGATACAAAGCCAA

GTACATGGTATGTTTCGTAATCGCTTACAATATACGCAGGAGTTTGTTTTTACAAGTACCGACAGC

TGGAGAAACTCACTCGCGAAGTCATACCAAGTGATGATAGTATACGAACCGCGAGACAAAAACGCC

TGAGGAGAAAGCTCTACTTTCTCACTCTCTCTATTCTCGTGTTAGTTGTGCCGATTGTTTGCGTCT

TCTTCGTCTTCAACCTCATTCTGGGTTGGCCGTGGTCGCTGCCGTTCGACCTCCATCGCATTCACG

CCAACATCAACTTTGTCTCGTTCACAACGACGGAAAGAATGCAAGTTACCGCGGTCCTTACCAATT

ACGTTCCAGTGGTGAGCTCGGCGGCCATTTTCATCACCTTCGGCACAACAGTTGAGGCATATAACC

AGTACCGTTTGGTTCTCGTTTTCCTCGGCTTTGGAAAGATTTGGCCCAAGCTGTATCAAGAATATG

ACCCCGATGACTCTGAGCCACCATCTGAGCTTAGCACGCAAACAAGTTCGAAATCGTGGTGGTCAT

CCATGACGAAGAACAGCAAAAGGGGTACAGTCACAAGGTATGAGACTCGCAAGCTTCCCAACATAA

CCATCACATTGGACTATCTGGATGCTAACGCACAATTGAATGAACAGCCAGAGCGGAAACGATTTC

TCTATCCTCCCTATTACTGAGAACATACCTCTCACGAACCGATCAACACCTGCACAGACAACGGGT

GAATGCTCCCATACCGTCCCATGTTCTCCATACGATGAGCTCAGAAAAGAATTCCCGGACCTGTAC

CCTACCCAGTCCTCACCCACAGCAAAAGCCAACCATAACGCCTGGCCCGATCTGTCAGACGACCCT

CCTGCTCCAGCACGCAGCCCATGGTATCTCCGACCGGATGCTTTCCATGTCCCGATCAAACTGCCG

ATGCCACTGTCACCCATTTACATACCATCCCTGCACGCGGATCAAGAAAAGAAGACGGAGGGAAAG

CAGAAACGAGCCAGTGTCAACAGGCTCTCTGCAGGCTACACCGCGCTCAGTGGTTCCTCCACAACA

ATCCTGCCGTCCTCGCAGTCACCCACCTGCCCTCTTCCCCCGCTTCCAACCCCCACATCCTCCTCA

AGATATTCACCACAGGAGGAGCGCCACCGGCCAACAAGAAAGAAACAAAAACAGAAAGAACCAGAG

ACGCAAAACAGCACAAGCAATGGGACACATAAGAAGAACTCATTGACGCCAGCAGCACCATGGAGT

CCCTTCGCTTCTACCCACACCCACACCACACTAGGAGTTGACACTCGCGTATGGTCGTCCCAGAAG

```
GATAACAACGGAAAAGACCATGCTGCTGCTCACCCACAAGACTCGCCAGTCTCCCATTTTTCTCGC

TTAACCTCAGGACCTCAGCTTTCACCGCCGCCATCGTACAGCGTCGCTGTCGGTGAATCCTCCTTT

GCTTCCGCTTCCGATCGATCGACGGGACTCGGAATCAGTTCGCCCGAACAACAACAGATCGGGACG

GCGTTGGGACACCCGCATCCGCTAAGGGCCCTCGGGGTGGTGAGGGTTGAGACGCATATAGCGAGC

GAGATTGAGGTTATTGGTGGTTCTGGACGTGATGATGGTAATGATGGTGATGGTGATCAACGGGGT

GGAAGGGAGTGA
```

CDS annotation:

```
join(1..643,700..1225,1302..2322)
```

Protein:

```
MNNTDSWASVANSTTSYTTTPEERLGPPAPYTDIGLQVNLFFRVFLGILGILIPLVPAKLLWINGE

FGATVHCMSTVTLNFFYVVNSLIWRDNNVKKWYAGYGWCDFHTYVFFAVETIFHTTLFDIMLGLAN

KIGNPRVTSLSPKEKKRKDRISALIIFGNPVLQVLLTYFVITQRYDISTLAGCNAIFDPNGVFFVF

FILPSPVFTLGAAGLAGVCFYKYRQLEKLTREVIPSDDSIRTARQKRLRRKLYFLTLSILVLVVPI

VCVFFVFNLILGWPWSLPFDLHRIHANINFVSFTTTERMQVTAVLTNYVPVVSSAAIFITFGTTVE

AYNQYRLVLVFLGFGKIWPKLYQEYDPDDSEPPSELSTQTSSKSWWSSMTKNSKRGTVTSQSGNDF

SILPITENIPLTNRSTPAQTTGECSHTVPCSPYDELRKEFPDLYPTQSSPTAKANHNAWPDLSDDP

PAPARSPWYLRPDAFHVPIKLPMPLSPIYIPSLHADQEKKTEGKQKRASVNRLSAGYTALSGSSTT

ILPSSQSPTCPLPPLPTPTSSSRYSPQEERHRPTRKKQKQKEPETQNSTSNGTHKKNSLTPAAPWS

PFASTHTHTTLGVDTRVWSSQKDNNGKDHAAAHPQDSPVSHFSRLTSGPQLSPPPSYSVAVGESSF

ASASDRSTGLGISSPEQQQIGTALGHPHPLRALGVVRVETHIASEIEVIGGSGRDDGNDGDGDQRG

GRE*
```

N. pannonica, mfa-1, a-factor pheromone

Gene sequence:

```
ATGCCTTCCACCGCTGCTTCCACCAAGGTCCCCCAGACCACCATGAACTTCAACGGCTACTGCGTT

GTTATGTAA
```

CDS annotation:

```
(1..75)
```

<u>Protein:</u>

```
MPSTAASTKVPQTTMNFNGYCVVM*
```

N. pannonica, ccg-4, α-factor pheromone

<u>Gene sequence:</u>

```
ATGAAGTTCACCCTCCCTCTTGTCATCTTCGCCGCCGTGGCCTCCGCCACTCCGGTCGCCCAGCCC
GCCGCCGAGGCCAACGCTCAGTGGTGCCAGCTCCACGGCCAGTCCTGCTGGAAGGTCAAGCGTGTC
GCCGAGGCCTTCGCCAACGCCATCCAGGGCATGGGTGGTCTCCCGCCCCGCGACGAGTCCGGCCAC
CAGCCCGCCCAGGTCGCCAAGCGCCAGGTCGACGAGCTCGCCGGCATCATCGCCCTCAGCCAGGAG
GACGTCAACGCCTACTACGACTCCCTCGGCCTCCACGAGAAGTTCGCCCCCTCCACCGAGGAGGAG
AAGAAGACCGAGAAGGTCGCCAAGCGCGAGGCCGAGGCCGAGGCGCAATGGTGCCGCATCCACGGC
CAGTCGTGCTGGAAGAAGCGTGAGGCGGAGGCTCAATGGTGCCGTATCCACGGCCAxxxxGCCGAG
CCCCAATGGTGCCGCATCCACGGTCAATCTTGCTGGAAGAAGCGTGACGCCGCTCCCGAGGCTGCT
CCCGAGGCTGAGGCCAACCCGCAATGGTGCCGCATCCACGGCCAGTCCTGCTGGAAGGCCAAGCGC
GCCGCCGAGGCCATCATGACCGCCATCCAGTCCGCCGAAGCCGAATCTGCTCTTCTCCTCCGTGAC
ACCACCTTCAGCCCCGTCGACCGCGTCGGCAAGCGCGATCCCCAGTGGTGCCGCATCCACGGCCAG
TCCTGCTGGAAGCGTGATGCCTCCCCCGAGACGGCTTGCAACGCCCCCGATGGTGCTTGCACCAAG
GCCACTCGTGACTTGCACGCCATGTACAACGTTGCTCGTGCCATCGTCACCGCTCACTCCGATGAG
AACTAG
```

<u>CDS annotation:</u>

```
(1..864)
```

<u>Protein:</u>

```
MKFTLPLVIFAAVASATPVAQPAAEANAQWCQLHGQSCWKVKRVAEAFANAIQGMGGLPPRDESGH
QPAQVAKRQVDELAGIIALSQEDVNAYYDSLGLHEKFAPSTEEEKKTEKVAKREAEAEAQWCRIHG
QSCWKKREAEAEAQWCRIHGxxxxxxxAEPQWCRIHGQSCWKKRDAAPEAAPEAEANPQWCRIHGQSC
WKAKRAAEAIMTAIQSAEAESALLLRDTTFSPVDRVGKRDPQWCRIHGQSCWKRDASPETACNAPD
```

```
GACTKATRDLHAMYNVARAIVTAHSDEN*
```

<u>Note:</u> This protein was encoded by a gene that was found on the ends of two scaffolds and thus could not be fully assembled. This is why the xxx is found in the middle of the gene and predicted protein.

N. africana, mfa-1, a-factor pheromone

<u>Gene sequence:</u>

```
ATGCCTTCCACCGCTGCTTCCACCAAGGTCCCTCAGACCACCATGAACTTCAACGGCTACTGCGTT
GTGATGTAA
```

<u>CDS annotation:</u>

```
(1..75)
```

<u>Protein:</u>

```
MPSTAASTKVPQTTMNFNGYCVVM*
```

N. africana, ccg-4, α-factor pheromone

<u>Gene sequence:</u>

```
ATGAAGTTTACCCTTCCTCTTGTCATCTTCGCCGCTGTGGCCTCCGCCACCCCGGTCGCCCAGCCC
AACGCCGAGGCCGAAGCCCAGTGGTGCCGCATCCACGGCCAGTCCTGCTGGAAAGTCAAGCGTGTT
GCCGATGCCTTCGCCAACGCCATCCAGGGTATGGGTGGTCTCCCGCCCCGCGACGAGTCCGGCCAC
CAGCCCGCCCAGGTCGCCAAGCGCCAGGTCGACGAGCTCGCCGGCATCATCGCCCTCACCCAGGAG
GACGTCAACGCCTACTACGACTCCCTCGGCCTGCACGAGAAGTTCGCCCCCTCCACCGAGGAGGAG
AAGAAGACCGAGAAGGTCGCCAAGCGCGAGGCCGAGGCCGAGGCGCAATGGTGCCGCATCCACGGC
CAGTCGTGCTGGAAGAAGCGTGAGGCCGAAGCTCAGTGGTGCCGCATCCACGGCCAGTCTTGCTGG
```

```
AAGAAGCGTGATGCTGCTCCCGAGGCTGCTCCTGAGGCCAACTCGCAATGGTGCCGCATTCGCGGT
CAGTCCTGCTGGAAGCGTGATGCCTCCCCCGAGGCGGCTTGCAACGCTCCCGATGGTGCTTGCACC
AAGGCCACCCGTGACTTGCACGCCATGTACAACGTGGCTCGTGCCATTGCCACTGCTCACTCCGAT
GAGAACTAG
```

CDS annotation:

```
(1..669)
```

Protein:

```
MKFTLPLVIFAAVASATPVAQPNAEAEAQWCRIHGQSCWKVKRVADAFANAIQGMGGLPPRDESGH
QPAQVAKRQVDELAGIIALTQEDVNAYYDSLGLHEKFAPSTEEEKKTEKVAKREAEAEAQWCRIHG
QSCWKKREAEAQWCRIHGQSCWKKRDAAPEAAPEANSQWCRIRGQSCWKRDASPEAACNAPDGACT
KATRDLHAMYNVARAIATAHSDEN*
```


H. bhutanensis, *MAT1-2-1*, MAT1-2-1 protein

Gene sequence:

```
ATGGAAAACTTCCTCATTGATCCCACCTTGATGGGCCCAGCTCATCTTGAGACTTCGGCTTTGGAA
ACTGCAGGTCTTGATCCATTGGTATACTTCACAGTCCAACAACAACAGATTCTCCAATCTGCATGG
GCAGCAGCAACAGTCCAAATATCACCATTTTCCAAAGTCGCAGCATTACATGCGAACATGGTACTA
GCCCTCAGCGAGGACTCCCAAAAGTCATTGCTGGCTGATTTCACGTACGAATATTTACTTCAAACC
AATTTTTAGTATACTGACATCATTTCTCCAGGAACGTGATAGGCGCCCCGGCCTTATTGGTTAGAG
ATTCTTCCGATCTCGATCGCTTTTTTATCGGATCCTTACAAGAATTTTCGTCCATCAACCAGTCGC
TCATCATGATGCCAGGATATGATCATTTCCTTTTGGTTTCGACTGGAGATGCTTCCATCCAAAGTG
GTCTCCCCTCTCCAATAAGCGTTCAAGGTCACGAATCCGACGACAGTAATAAAGATGACGTCAAAC
AAAAACTACCTCGCCCGCCTAATGCGTACATATTGTACCGTAAAGAGCGCCACCATTCCGTCAAAG
ACGAGTTTCCGGGGATCTGTAACAACGAAATTTGTGAGTCCTAATTATCATACAACCCTCACTGAT
ATAAGGCTTACTAACATAGCTATTAAGCGCGAATTCTTGGCCGCCGATGGAAGGAAGAAAGCGAGA
CAGTTCGGGCGTTCTACAAGAGCAATCTGAAAACTACAAACAAAACTTCATGAACACCCACCCAG
ACTATCAATACAGACCGCGCAAGGCAGGAGAAAAGAAGAAGAGGAACAGGCGTGTTCAACCCAAAG
```

```
ACTCCGAGAACACTGGGTTACAATCACCCGTTTCTGCTAGGGGAACCCCAAAGGTGTTGTTCGCTG

AGAGCCCTATTTAA
```

CDS annotation:

```
join(1..242,296..627,688..938)
```

Protein:

```
MENFLIDPTLMGPAHLETSALETAGLDPLVYFTVQQQQILQSAWAAATVQISPFSKVAALHANMVL

ALSEDSQKSLLADFTNVIGAPALLVRDSSDLDRFFIGSLQEFSSINQSLIMMPGYDHFLLVSTGDA

SIQSGLPSPISVQGHESDDSNKDDVKQKLPRPPNAYILYRKERHHSVKDEFPGICNNEISRILGRR

WKEESETVRAFYKEQSENYKQNFMNTHPDYQYRPRKAGEKKKRNRRVQPKDSENTGLQSPVSARGT

PKVLFAESPI*
```

H. bhutanensis, *MAT1-2-7*, MAT1-2-7 protein

Gene sequence:

```
ATGGACATTGACGCCGTGCGACACCTCCAGCACCTTTGGGTAACCTGTCGCGAAAAACCAAAATGG

CAGATGCTGGCGTCTCTCATTCTTATCATTCACACCATCCAGACTCAGGAAGGAGTACCCCTCATC

CATCCTTGCGCATTGATACCATCAGCCGCCTCTTTGCTCGAAGTTATCCCCCCAAATCGATATCTC

CAGAACTGCTTCCAACAGCCACCTTCCTTCACAATGGACCTCCCTCTTCTCTTCAAGAGACTAGCT

TTTGAGAAAGGCTTCATGGGGATTTCCGGCCGATACATGTACCGATACTATGTCTCTGTCTACATG

AACGCCATGCTTGATCTCCCAGAAGATGCAGTTTCATTGCATGGAGTCAGTGCCTTTGCACATCAA

ATTTCACGCCTCACATCCCGAAGCTCTCGCGCATACTTCATCGAACTGGGCAACGGAGTTCTCTGG

GCATAA
```

CDS annotation:

```
(1..468)
```

Protein:

```
MDIDAVRHLQHLWVTCREKPKWQMLASLILIIHTIQTQEGVPLIHPCALIPSAASLLEVIPPNRYL

QNCFQQPPSFTMDLPLLFKRLAFEKGFMGISGRYMYRYYVSVYMNAMLDLPEDAVSLHGVSAFAHQ

ISRLTSRSSRAYFIELGNGVLWA*
```

H. bhutanensis, a-factor pheromone

<u>Gene sequence:</u>

```
ATGGCCGCTATCAAGAACATCACCTCCTCCAAGAACGCCGCCCGCGGCGTCGACCAGTCCAACCCG
TGCAACGTCATGCGTGGTGTTGACCAGTCCAACCCGTGCGCTGTCATGCGCGGTGTTGATCAGTCC
AACCCCTGCACCGTCATGTAA
```

<u>CDS annotation:</u>

```
(1..153)
```

<u>Protein:</u>

```
MAAIKNITSSKNAARGVDQSNPCNVMRGVDQSNPCAVMRGVDQSNPCTVM*
```


H. bhutanensis, α-factor pheromone

<u>Gene sequence:</u>

```
ATGAAGTTCTCTACCATCCTGGCTGTTGCCTCTGGTCTCGCTGCCGTCGAGGCCGCTGCCATCACC
CCCGTCGAGGCCCGCGATACCGCCCCTACTGTTGAAGCTCGCGGCAAGTTCTTGAAGAACTCCAAC
GGCGGTCTCCCCGGTGAACTCCTCCGCCGCTTCTTGGACGACTCCAACGGTGGTCTTCCCGGTGAG
CTTCTTCGCCGCGAGGAGCCTACTCTTGAGGCTCGCGGCAAGTTCCTGAAGAACTCCAACGGTGGT
CTCCCCGGTGAGCTTCTCCGTCGTGAGGAGTCCATCGAGGCCCGTGATGAGGCTACTCTTGAGGCT
CGTGGCAAGTTCCTGAAGAACTCCAACGGTGGTCTTCCTGGTGAGCTTCTCCGCCGCAAGTTTCTC
CAGAACTCCAACGGTGGTCTTCCCGGTGAGCTCCTCCGCCGTGAGGAGGCTACCCTTGAGGCCCGC
GGTAAGTTCTTGAAGAACTCCAACGGTGGTCTTCCCGGTGAGCTTCTCCGCCGTGAGGAGCCTACC
CTTGAGGCTCGTGGCAAGTTCCTGAAGAACTCCAACGGTGGCCTTCCTGGTGAGCTTCTCCGCCGC
TTCTTGGACGACTCTAACGGTGGCCTTCCCGGTGAGCTTCTTCGCCGTGAGGAGCCTACCCTTGAG
GCTCGTGGCAAGTTCCTGAAGAACTCCAACGGTGGTCTTCCCGGCGAGCTTCTCCGTCGCGAGGAG
TCCATCGAGGCTCGCGATGAGGCTACCCTTGAGGCTCGCGGCAAGTTCCTGAAGAACTCCAACGGT
GGTCTTCCTGGTGAGCTTCTCCGCCGCGAGGAGCCTACTCTTGAGGCCCGCGGTAAGTTCTTGAAG
AACTCCAACGGCGGTCTCCCCGGTGAGCTTCTGTAA
```

<u>CDS annotation:</u>

(1..894)

<u>Protein:</u>

MKFSTILAVASGLAAVEAAAITPVEARDTAPTVEARGKFLKNSNGGLPGELLRRFLDDSNGGLPGE

LLRREEPTLEARGKFLKNSNGGLPGELLRREESIEARDEATLEARGKFLKNSNGGLPGELLRRKFL

QNSNGGLPGELLRREEATLEARGKFLKNSNGGLPGELLRREEPTLEARGKFLKNSNGGLPGELLRR

FLDDSNGGLPGELLRREEPTLEARGKFLKNSNGGLPGELLRREESIEARDEATLEARGKFLKNSNG

GLPGELLRREEPTLEARGKFLKNSNGGLPGELL*

H. bhutanensis, *pre1*, a-factor pheromone receptor

<u>Gene sequence:</u>

ATGAGTAATTTCTCAAACAGCAATCCGGCCCTCCGCATAGTCGATGGCATGTCTGTTATCGACTTC

TCATACCTCATGACGAGCCCATCAAACCCCAACCAAATAATTCCCAACAGCGATGGCGTCGAAGCT

TACCCCGGCCATGTATATACTGATCGAATCCTTCGACTCACACTCGGCCCTCTTGCCATTCTCCTT

TACTTTATGCCACTTCGTATCCTCTACAGACACAATGAGTTCTGCGCCTTCATCATGGTGTTTTGC

TGGTTGATAAACACCGTGTTTTTCACTCTTAGCGCTATACTGTGGCCGTCGCATGATACCAGTAAT

TGGTGGTCTGGAAAAGGGTATTGTGATGTTCATTTAACATTCTACGCGGCTGTTCCTACCATGGCC

ATGTCTGCTCTCTACATCATGTTAGCTACCCTTTCTCGCTCCATAACACTCGCCAGTGAGGGTCCG

TTATTACGGGCTGAAGCTCGCCGTGTAAAGCTCAAACAAGGCCTCTTCATTTTCACACCCGCTATC

ATCGTTGGCCTCTTATCAAATCTCGCTGCCTCCAACCGGTTCGTTTTATATCCTATAGCCGGCTGC

GTACGCCTCTGGGATACGTCATGGCCTACCCTTGCCGTCAACATCATTCAAGAGATATTCTGCCTC

ATGGCTGGGTTTACAGCTGGTAAGTAACAAAGATTCCGCACGTCAACTGCTACAAATCTTACTGAC

ACTACTTTACTAGTCAAACTTTTGTGGACTTATTACAGGTCTCAACGGAGAAACAAATCTGCCACC

GCCCTGATGAACAATCCGCTAGCCCGCCAATCCAGTAACCGTGCCCGTCGTCGTCTAGTCATGCTC

GCAGTAACCATTCTAGTACCCTACGTCCCCATATCGACATACATTATCATCACAAACATCAAGGTT

CTTATACCCTTGAATGCGTATTCCTTCGCCGGGCTCCATAAGAACATAGATTTAGGCTTTGGTGAG

```
ATTTCATGGCAGGCAATATATTTCATGACAGGCGAATGGGTAAATAACCGTGCTGTATTGAACACA
TGGATTCCAATAGCTGCAGTTATCCCATGTTTCGGGTTCTACGGGCTCTCAACCGAAGCTATAAAC
TCCTATCGGAGAGGTCTCCTCAACCTCGGTCTTGGCCCCCGGTATCCAATTCTCTACACGGAATAC
ACACCCGGTCGTCGCACATCATCAACGCCCTCATGGCTCGCCTCATTCGTTTCTTCGCGGCCTGGA
GAGATCTCCAAGAATAAACCATCTGAATTAAGTTCGGGGAGCGCATGCGCATCATTACAAGGACTT
GAAAGCCCAGCAGTTGCCCACATTAAAACAGAAAGTGGTAGACCAATGGTATGGGACGGTCTTTCT
CGAGCAACAGGTGGAGAGGCTGGTGCTGTAACCACTGATTGTTATGGCGAGGGGAGTCTACAGAGA
GTAGGCACCGGTGACGGGAACGTGACGGTGATGACGGAGATTGACCTGAGAAGTACCCGAGGCGAG
CGGGCTCAAGAAGAAGTATAG
```

CDS annotation:

```
join(1..679,740..1539)
```

Protein:

```
MSNFSNSNPALRIVDGMSVIDFSYLMTSPSNPNQIIPNSDGVEAYPGHVYTDRILRLTLGPLAILL
YFMPLRILYRHNEFCAFIMVFCWLINTVFFTLSAILWPSHDTSNWWSGKGYCDVHLTFYAAVPTMA
MSALYIMLATLSRSITLASEGPLLRAEARRVKLKQGLFIFTPAIIVGLLSNLAASNRFVLYPIAGC
VRLWDTSWPTLAVNIIQEIFCLMAGFTAVKLLWTYYRSQRRNKSATALMNNPLARQSSNRARRRLV
MLAVTILVPYVPISTYIIITNIKVLIPLNAYSFAGLHKNIDLGFGEISWQAIYFMTGEWVNNRAVL
NTWIPIAAVIPCFGFYGLSTEAINSYRRGLLNLGLGPRYPILYTEYTPGRRTSSTPSWLASFVSSR
PGEISKNKPSELSSGSACASLQGLESPAVAHIKTESGRPMVWDGLSRATGGEAGAVTTDCYGEGSL
QRVGTGDGNVTVMTEIDLRSTRGERAQEEV*
```

H. bhutanensis, *pre2*, α-factor pheromone receptor

Gene sequence:

```
ATGACCGATTCACTTTTATTTCTTCATACAATATCCACTTTTCAGAGCCACTTGATGGCTCGTGAT
GATAGTAACCAGACCATCGACTTCGACCCGAAAAGTGTACCGTTCAACATCACCACTCCAAGTGGT
GGAGTCGTGACAGTAACTATGCATGTCGTCGAATACTACTCTCGGTATCTGGTTAACTACTCTATA
TCACGCGGGCTACAAACTGGAATGTTAGCCATGACATTAATGGCTTTGATATTTTTCAGCCCCAAG
```

```
GACCGCATCCTCAAAAAGAACATCGTCATTCAAATCATCACTTTATGCGTTGCCATCATCAAAAGC

GTCCTCCTACTCAGATGGGCTACATCGGGGTGGGGATCGATGTACAGCATCATCACCGGAGATTAC

TCCAACCTGAAGGCTTCAGACTCCAACAACCAAATCGTTCTATCAGTCGCTAGCATTCTACTCATT

CTTTTTGTTGAACTGATGCTGCTGCTACACAGCTGGACTATCTTTCGCATGTGGAGGCCGACCATA

AAGTACATTTGTTTGGTTTTATCAGTGTTCTTAGGAACTTCCACAGTCGCTATGCGAACAGTATAT

CAGGCAATAGGACTGAAATCTGTGTTAGATGTGGCGGCTGACAGGGAATCATCAGAGTCTACCGAC

TTGTTACTCGTAACTGTATTGTTTTTCTCAAACTTGGCTCTCGCTATTTCAACCGTTTGGTTTACA

GCGATCTTCAATGTCAAATTAATTAGTCATATGTGGAAGACTAAACGGCTGCTACCCAGTCGTAAA

GGCATGACGCCGATGGAGATTTTAACTGCAACTAATGGTTTTCTCATGATTGCACCATGTAAGTGT

AATTGAATTTGTCTTGTGTCATTATAGATACTAATCTTCCTTTCTTCAGTGATCTTTGTTTGTCTG

TCCAGCTCGAAGCTCTCACACACAGTAGACTTTGCTACATGGACCACCACTGCCATTCCTATGATT

CTGCCATTTGGGTCCCTCATTGCTCAAAGACTTATCAGCGGCCAAAAAGTCTCTTCACTTGTTCAC

ATCTCAGCCGTGTCAAGCAACAGCAGCAATACCCCAACACTTCCATTGTTTAACGGAGGACAAGAC

CTTACCATGACTGCAGCAACTAACAAGACAATGGTCGAGACCGGGGCGACTGGCGAGTTTCCAGAT

GATCGCGATGATTGTGTCAATATTCACAGAGATGTTGACCAAAGAAGTGAGCGTCGTGTAGCTCAA

AATTGA
```

CDS annotation:

```
join(1..850,908..1260)
```

Protein:

```
MTDSLLFLHTISTFQSHLMARDDSNQTIDFDPKSVPFNITTPSGGVVTVTMHVVEYYSRYLVNYSI

SRGLQTGMLAMTLMALIFFSPKDRILKKNIVIQIITLCVAIIKSVLLLRWATSGWGSMYSIITGDY

SNLKASDSNNQIVLSVASILLILFVELMLLLHSWTIFRMWRPTIKYICLVLSVFLGTSTVAMRTVY

QAIGLKSVLDVAADRESSESTDLLLVTVLFFSNLALAISTVWFTAIFNVKLISHMWKTKRLLPSRK

GMTPMEILTATNGFLMIAPLIFVCLSSSKLSHTVDFATWTTTAIPMILPFGSLIAQRLISGQKVSS

LVHISAVSSNSSNTPTLPLFNGGQDLTMTAATNKTMVETGATGEFPDDRDDCVNIHRDVDQRSERR

VAQN*
```

H. omanensis, a-factor pheromone

<u>Gene sequence:</u>

```
ATGGCCGCTATCAAGAACACCACCACCTCCAAGAACGCCGCCCGCGGCGTTGACCAGTCCAACCCC
TGCGCCGTCATGCGCGGCGTTGACCAGTCCAACCCCTGCGCTGTCATGCGCGGCGTCGACCAGTCC
AACCCCTGCACTGTCATGCGCGGTGTTGACCAGTCCAACCCGTGCACCCTCATGTAA
```

<u>CDS annotation:</u>

```
(1..189)
```

<u>Protein:</u>

```
MAAIKNTTTSKNAARGVDQSNPCAVMRGVDQSNPCAVMRGVDQSNPCTVMRGVDQSNPCTLM*
```

H. omanensis, α-factor pheromone

<u>Gene sequence:</u>

```
ATGAAGTTCTCTACCATCCTGGCTGTCGCCTCTGGTCTTGCCGCCGTCGAGGCCGCCGCCATCGCC
CCCGTTGAGGCCCGCGATGCCACCCCCGCTGTTGAGGGTCGTCACTTCCTGGACGACTCCAACGGT
GGTCTCCCCGGTGAGCTTCTCCGCCGCGAGGAGTCTACCGTCGAGGCTCGCCACTTCTTGGAAAAC
TCTAACGCTGGCCTCCCCGGTGAGCTTCTTCGCCGCTTCTTGGAGGACTCCAACGGTGGTCTTCCT
GGTGAGCTTCTTCGCCGCTTCTTGGAGGACTCTAACGGTGGTCTTCCCGGTGAGCTTCTTCGCCGT
GAGGAGCCTACCGTCGAGGCTCGCCACTTTCTTCGTCGCTTCTTGGAGGACTCCAACGGTGGTCTT
CCTGGTGAGCTTCTCCGTCGCTTCTTGGAGGACTCCAACGGTGGTCTTCCCGGTGAGCTTCTCCGC
CGCGAGGAGCCTGCCGTCGAGGCTCGTCACTTCCTGAAGTACTCCAACGCCGGTCTCCCCGGTGAG
CTTCTCCGCCGTGAGGAGCCTGCCGTTGAGGCTCGCCACTTCCTGAAGTACTCCAACGCTGGTCTC
CCCGGTGAGCTTCTGTAA
```

<u>CDS annotation:</u>

```
(1..612)
```

Protein:

```
MKFSTILAVASGLAAVEAAAIAPVEARDATPAVEGRHFLDDSNGGLPGELLRREESTVEARHFLEN
SNAGLPGELLRRFLEDSNGGLPGELLRRFLEDSNGGLPGELLRREEPTVEARHFLRRFLEDSNGGL
PGELLRRFLEDSNGGLPGELLRREEPAVEARHFLKYSNAGLPGELLRREEPAVEARHFLKYSNAGL
PGELL*
```

H. omanensis, *pre1*, a-factor pheromone receptor

Gene sequence:

```
ATGAGTGATTTCTCAGGCAGTAATCCGGCTCTTCGCATGGTCGATGGCATGTCTGTCATCGACTTC
ACATACCTCATGACAAGCCCATCAGACCCCAATCAGATAATTCCCAACAGCGATGGCGTTGAAGCT
TACCCCGGCAATGTATATACTGATCGAATCCTTCGACTTACACTCGGCCCTCTTGCCATTCTTCTT
TACTTTATGCCACTTCGTATCCTCTACAGACACAATGAGTTCTGCGCCTTCGTCATGGTGTTTTGC
TGGTTGATAAACACTGTGTTTTTCATTCTTAACGCTATAATGTGGCCGTCACATGATACCAGAAAT
TGGTGGTCTGGAAAAGGGTATTGTGATGTTCATTTAACATTCTATACGGCTGTTCCTACCATGGCC
ATGTCTGCTCTCTACATCATGTTAGCTACCCTTTCTCGCTCCATAACACTCGTCGCCGAGGGTCCG
TTATTACGGGCTGAAGCTCGTCGTGTAAAGCTCAAACAAGGCCTCTTCATTTTCACACCCGGAATC
ATCATTGGTCTCTTATCAAATCTCGCTGCCTCCAATCGGTTCGTTTTATATCCCATAGCCGGCTGC
GTGCGCCTCTGGGATACGTCATGGCCTACCCTTGCCGTCAACATCATTCAAGAGATATTCTGCCTC
ATGGCTGGGTTTACAGCTGGTAAGTGATACAGATCCCGCAAATCAACTGCTATAAATCTTACTGAC
ACTACTTTACTAGTCAAACTTTTATGGACTTATTACAGGTCCCAACGGAGAAACAAATCTGCCACT
GCCCTAATGAACAATCCGCTAGCCCGCCAGTCCAGCAACCGTGCTCGTCGTCGTCTCGTCATGCTC
GCAGTAACCATTCTAGTGCCCTACGTCCCCATATCCACATACGTCATCATCACAAACATCAAGGTC
CTTATACCCATGAATACGTATTCCTTCGCCGAGCTCCATAAAGACAGAGATATAGGCTTTAGTGAG
GTTTCATGGCAGACAATATATTTCATGACAGGCGAATGGGTAAATAACCGTGCTGTATTGAACACA
TGGATCCCAGTAGCTGCAGTTGTCCCATGTTTCGGGTTCTACGGGCTCTCAACCGAAGCTATAAAC
TCCTATCGGAGAGGTCTCCTCAACCTCGGTCTTGGCCGGCGGTATCCAATTCTCTACACGGAATAC
ACACCCGGTCGTCACACATCATCAACACCCTCATGGCTCGCCTCATTCGTTTCTTCGCGGCCTGGA
```

```
GAAATCTCTAAGAATCAACCATCGGAATTAAGTTCAGGGAGCGTATGCGCATCATTACAAGGACTT

GAAAGCCCAGCAGTTGCCCATGTTAGAACGGAAAATAGTAGACCAATGGTATGGGATGGTCTTTCT

CGGGCGACAGGTGGGGAGGCTGGTGCTGTGACCACTGATTGTTATGGTGAGGGCAGTCTACAGAGA

GTAGGCACCGGCGACGGGAACGTGACGGTGATGACGGAGATTGATCTGAGAAGTACCCGAGGCGGG

CGGGCTCAAGAAGAAGTATAG
```

CDS annotation:

```
join(1..679,815..1539)
```

Protein:

```
MSDFSGSNPALRMVDGMSVIDFTYLMTSPSDPNQIIPNSDGVEAYPGNVYTDRILRLTLGPLAILL

YFMPLRILYRHNEFCAFVMVFCWLINTVFFILNAIMWPSHDTRNWWSGKGYCDVHLTFYTAVPTMA

MSALYIMLATLSRSITLVAEGPLLRAEARRVKLKQGLFIFTPGIIIGLLSNLAASNRFVLYPIAGC

VRLWDTSWPTLAVNIIQEIFCLMAGFTAARQSSNRARRRLVMLAVTILVPYVPISTYVIITNIKVL

IPMNTYSFAELHKDRDIGFSEVSWQTIYFMTGEWVNNRAVLNTWIPVAAVVPCFGFYGLSTEAINS

YRRGLLNLGLGRRYPILYTEYTPGRHTSSTPSWLASFVSSRPGEISKNQPSELSSGSVCASLQGLE

SPAVAHVRTENSRPMVWDGLSRATGGEAGAVTTDCYGEGSLQRVGTGDGNVTVMTEIDLRSTRGGR

AQEEV*
```

H. omanensis, pre2, α-factor pheromone receptor

Gene sequence:

```
ATGAGCGATTCACCTTTTTTTTCTTCATACGATATCCACTTTTCAGAGCCACTTGATGGCTCGCGAT

GATAGTGACCAGACCATAGGCTTCGACCCGAAAAGTGTACCGTTCAACGTCACCACTTCAAGCGGT

GGAGTCGTGAAAGTAACTATGCATGTCGTCGAATCCTACTCTCGGTATCTGGTCAACTACTCTATA

TCACGCGGGCTACAAACTGGAATGTTAGCCATGACATTAATGGCCTTGATATTTTTCAGCCCCAAG

GACCGCATCCTCAAAAAGAACATCGTCATACAAATCATCACTTTATGCGTTGCCATCGTCAAAAGC

GTCCTCCTACTCAGATGGGCTACATCGGGATGGGGATCGATGTACAGCATCATCACTGGAGATTAC

TCCAACTTGAAGGCTGCAGACTCCAATAACCAAATCGTTCTATCAGTCGCTAGCATTCTACTCATT

GTTTTTGTTGAACTAATGCTGCTACTACACAGTTGGACTATCTTTCGCATGTGGAGACCGATCATC
```

```
AAGTACATTTGTTTGGCTTTATCAGTGTTGTTGGGATCTTCCACAGTCGCTATGCGAACAGTATAT
CAGGCAATAGGACTGAAATCTGTGTTAGATGTGGCGGCTGACAGGGCATCGTCAGAGTCTACCGAC
ATGTTACTTGTAACTGTATTCTTTTTTTCAAACTTGACTCTCGCTATTTCAACTGTTTGGTTCACA
GCCATATTCAATGTCAAGCTGATTAGTCATATGTGGAAGACTAAACGGCTGCTACCCAGTCGTAAA
GGCATGACGCCGATGGAGATTTTAACTGTAACTAATGGTTTTCTCATGATTGCACCAAGTAAGTGA
AATTCATCCTGTCTTGCGTTATTATACAATACTAATCTTCTTTTGTTCAGTGGTATTTGTTTGTCT
GTCCAGCTCGAAGCTCTCACACGTAGCAGACTTTGCTACATGGACGACCACTGCCATTCCTATGAT
TCTGCCATTTGGGTCCCTCATTGCTCAAAGACTCATCACCAGCCAAAAGTCTCTTCACTTGTTCA
CATCTCAGCTGTGTCAAGCAACAGCAGCAATACACCAACACTTCCATTGTTTAACGGGGGACAAGA
CCTTACCATGACTGCAGCAACCAACAAGACAATGGTCGAAACCGGGGCGACTGGCGAGTTTCCAGA
TGATCGCGATGATTGTGTCAATGTTCACAGAGATGTTGACCAAAGAAGCGAGCGCCGTTTAGCTCA
AAGCTGA
```

CDS annotation:

```
join(1..850,909..1261)
```

Protein:

```
MSDSPFFLHTISTFQSHLMARDDSDQTIGFDPKSVPFNVTTSSGGVVKVTMHVVESYSRYLVNYSI
SRGLQTGMLAMTLMALIFFSPKDRILKKNIVIQIITLCVAIVKSVLLLRWATSGWGSMYSIITGDY
SNLKAADSNNQIVLSVASILLIVFVELMLLLHSWTIFRMWRPIIKYICLALSVLLGSSTVAMRTVY
QAIGLKSVLDVAADRASSESTDMLLVTVFFFSNLTLAISTVWFTAIFNVKLISHMWKTKRLLPSRK
GMTPMEILTVTNGFLMIAPMVFVCLSSSKLSHVADFATWTTTAIPMILPFGSLIAQRLITSQKVSS
LVHISAVSSNSSNTPTLPLFNGGQDLTMTAATNKTMVETGATGEFPDDRDDCVNVHRDVDQRSERR
LAQS*
```

H. moniliformis, a-factor pheromone

Gene sequence:

```
ATGCCTTCCATCAAGAACCACACTCCGTCTACCAAGACCTCTGGCAATGAGACCATCCAACCCCCC
ACTAGCAACGCCGGGCGTGGAGCGATTCAGTCTCCCATCAACCCGATCACCCGCGGCGTCACCCAG
GCTCCGCCGTGCAATGTGATGCGTGGAGTTACTCAAGCCCCCCCTTGCAATGTCATGTAA
```

<u>CDS annotation:</u>

```
(1..192)
```

<u>Protein:</u>

```
MPSIKNHTPSTKTSGNETIQPPTSNAGRGAIQSPINPITRGVTQAPPCNVMRGVTQAPPCNVM*
```

H. moniliformis, α-factor pheromone

<u>Gene sequence:</u>

```
ATGAAGTTCTCCACCATCTTGGCTGTTGCCTCCGTTCTTGCCGCCGTTGAGGCCGCTGCCATTGAG
CCCCGCCGCGACATCTTGAAGGACGCCAACGGTGGTCTTCCTGGCGAGCTTTTCCGCCGCGAGGAG
TCCGTCATTGAGCCCCGCCGCAAGATCCTCAAGGACGCCAACGGTGGTCTTCCCGGTGAGCTCTTC
CGCCGCGAGGAGGCTGCTGTTCTCGAGCCCCGCCGCAAGATCCTCAAGGACGCCTGGGGTGGTCTC
CCTGGCGAGCTTTTCCGCCGTGAGGAGTCTGTCATTGAACCCCGCCGCAAGATCCTCAAGGACGCC
AACGGTGGTCTCCCTGGCGAGCTCTTCCGCCGCGAGGAGGCTGCCGTTCTCGAGCCCCGCCGCAAG
ATCCTCAAGGACGCCTGGGGTGGTCTCCCTGGCGAGCTTTTCCGCCGTGAGGAGTCTGTCATTGAA
CCCCGCCGCAAGATCCTCAAGGACGCCAACGGTGGTCTTCCCGGTGAGCTCTTCTAA
```

<u>CDS annotation:</u>

```
(1..519)
```

<u>Protein:</u>

```
MKFSTILAVASVLAAVEAAAIEPRRDILKDANGGLPGELFRREESVIEPRRKILKDANGGLPGELF
RREEAAVLEPRRKILKDAWGGLPGELFRREESVIEPRRKILKDANGGLPGELFRREEAAVLEPRRK
ILKDAWGGLPGELFRREESVIEPRRKILKDANGGLPGELF*
```

H. moniliformis, *pre1*, a-factor pheromone receptor

<u>Gene sequence:</u>

```
ATGAGCAATCTCTCAGCCAACAACCCGGCCTTCCGTACGGTGAATGGAATGTCTATTCTCGACTTT

TCATATCTCATGACGGACGCAGCAAACCCAAACTACATAACAATGAACGCCGACGGTGTCCCTGCT

TATCCCGGCCATTTGTATACTGATCGAATACTCCGACTTATATTCGGGCCTCTTGTCATACTCCTT

TACCTTATCCCACTCCATATCCTCCACAGACACGATGAGTTCTGCGCCTTCATAATGGTTTCTTGC

TTTGTGGTATACGCCATATTTTTCACCCTCAACGCTATAATGTGGCCGTCACAGGACACTAGCAAT

TGGTGGTCTGGTATCGGGTATTGCGACATCCAGTTAACATTCTATACTGCCGTTCCTACAGTAGGT

ATGGCGGCACTTTACATGATGTTGGCTGCTCTGTCTCGTTCTATAACACTCGCCACAGAGGGCCCT

TTGCTACGGGCTCAAGCTCGCCGGTTAAAACTCAAGCAAGGACTCTTCATTTTCACACCTGGTGTC

TTTGTTGGCCTTCTATCAAACCTCGCTGCCTACAATCGGTTCGCTATATATCCTACCGCTGGCTGC

ATGCGTATCTCGGATGCATCATGGCCTGTCTTTGTCGTCAACATTATCCAAGAGTTATTTTGCCTC

ATGGCCGGGTTCACAGCTGGTAAGTTACACAGAACCCGCATGCATGTCAACCACCACAAATCTTAC

TGACATTTATTTCACTAGTCAAAATTTACTGGACTTATTTCAAGTCTCAACGAAGGAACAGCTCAG

CTACAGCCCTGATGAATAATCCTCTAGCCCGCCAATCCAGTAACCGGGCTCGTCGTCGTCTCTACA

TGCTCGCAGTAACCATTCTGGTGCCCTACGTCCCTCTGTGCACATACCTGATCACCATAAACATTA

AAGCTCTTGTACCCATGAATTCATATTCCTTCAAACAGCTCCATGCGCCCAAAGATTTAGGCTTTG

GTGAAGTTTCATGGCAGACCATCTATTTCATGGCAGGCGAATGGGTGAATACCCGTACTGCATTCT

ACACATGGGTTCCAATATTTGCAATTTTCCCATGTTTCGGGTTCTACGGTCTTTCAACCGAAGCTA

TAAACTCTTATCGGAGAGGTCTACTTAAACTCGGCCTTGGCTCCCGGTATCCAATTCTTTACACGG

AATACACACCCGGTCGTCGCGCGTTATCAAATCCTTCATGGCTTGCCTCATTTGTCACATCACAGC

CTAGAGATGTTCCCAAGACTCAACAATGTGAATTTAGTTCTGGAAGTCCTCGTACGTCACTCCAAG

GACTTGCGACCCCAGGAGTTGCTCATGTCAGACCAGATGATGACAGACCAATGGTATGGGACGGTC

TTGCTCGAGCAACAGGTGGTGGGGCTGGTGCAGTAACTACTGATTGTTATGGTGAGGGTAGTATAC

AGAGAGTAAGCACCGGTGACGGGAACGTCAAGGTGATGACGGAGATTGACCTGAGAAGCACCCGAG

GCGGACGGGATCAAGAAGAAGTATAG
```

CDS annotation:

```
join(1..679,745..1544)
```

Protein:

MSNLSANNPAFRTVNGMSILDFSYLMTDAANPNYITMNADGVPAYPGHLYTDRILRLIFGPLVILL

YLIPLHILHRHDEFCAFIMVSCFVVYAIFFTLNAIMWPSQDTSNWWSGIGYCDIQLTFYTAVPTVG

MAALYMMLAALSRSITLATEGPLLRAQARRLKLKQGLFIFTPGVFVGLLSNLAAYNRFAIYPTAGC

MRISDASWPVFVVNIIQELFCLMAGFTAVKIYWTYFKSQRRNSSATALMNNPLARQSSNRARRRLY

MLAVTILVPYVPLCTYLITINIKALVPMNSYSFKQLHAPKDLGFGEVSWQTIYFMAGEWVNTRTAF

YTWVPIFAIFPCFGFYGLSTEAINSYRRGLLKLGLGSRYPILYTEYTPGRRALSNPSWLASFVTSQ

PRDVPKTQQCEFSSGSPRTSLQGLATPGVAHVRPDDDRPMVWDGLARATGGGAGAVTTDCYGEGSI

QRVSTGDGNVKVMTEIDLRSTRGGRDQEEV*

H. moniliformis, *pre2*, α-factor pheromone receptor

Gene sequence:

ATGACTTATCCACCTTTCTTTGATCATACTATATCCACTTTCCACAGCTACTTAATGGCTCGTGAT

GAGAGTAACCAGACCATAGACTTCGATCCCAAAGATATAACATTCAACATCACCACCACAGAGGGT

GAAGTTCTGACCATCAATATGTATACCATCGAATACTACTCTCGATATCTGGTCAACTACTCTATA

TCACGCGGGCTCCAAACCGGAATGTTAGCCATGACATTGATGGCTTTGATATTTTTCAGTCCAAAA

GACCGCATCCTCAAGAAAAACATCGTTATACAAATCGTTACCCTGTGCATTGCTGTCGTCAAGAGC

GTCTTTCTGCTCAGATGGGCTACGTCGGGGTGGGGTTCAATGTACAGCATCATCACTGGAGATTAC

TCGTATCTCAAAGCTTCAGATTCCAACAACCAAATTGTTCTGTCAGTCGCTGGCATCTTTTTGGTT

CTTTTCATCGAGATAATGTTGTTGCTTCATGGCTGGACTATCTTTCGCATGTGGAAGCCCATCATC

AAGTACATCTGTTTGGTTCTGTCGGTGTTCTTCGCCATTGCAACAGTCGTTATGCGAGCTGTATAT

CAAGTAATAGGACTGAAATATCTTTTGACCACGGCGGCTGGCAAGCTGATGTCACAGGCTTCCGAC

GATTTAGTCGTGGTTTCATTGTTTTTTTGCAAATTTGGCTCTCGTTATTTCAACCATTTGGTTTACA

GCAATCTTTAATGTCAAATTGGTTAGCCACATGTGGAAGACTAAACGAATATTGCCTACCCGTAAA

GGTATGACACCGATGGAGATTTTAACCGTCACCAATGGTTTGCTCATGATTGTACCATGTAAGTCT

CAATTTCTCGTCTTCTGCCACGTTAGATACTAATCCTTCCTTGTTCAGTTGTATTTGTTTGTCTGT

CAAGTACAAAGCTTTCACATCAAATAGACTTTTCTACATGGACCATAACTGTAATCCCCATGATTT

TGCCATTTGGATCCCTCATTACCCAGAGGCTTATCGGCAACCAAAAAGTCTCTTCACTTGCTCACA

```
TATCAGTTGCGTCGAGCAGCACCAGCAACACACCAACGCTTCCATTGTTTAATGGGGGACAGGACC

TGACCATGACTCCAGCAACCAACAAGACCATGGTAGAAACTGGGGCGGCTGGCAAATTCCCGAATG

ACTGCGATAATTGTGTCAATGTTCACAGAGATATTGACCAGAGAAGCGAGCGTCGGTTGGCTCACA

ATTGA
```

CDS annotation:

```
join(1..850,1063..1259)
```

Protein:

```
MTYPPFFDHTISTFHSYLMARDESNQTIDFDPKDITFNITTTEGEVLTINMYTIEYYSRYLVNYSI

SRGLQTGMLAMTLMALIFFSPKDRILKKNIVIQIVTLCIAVVKSVFLLRWATSGWGSMYSIITGDY

SYLKASDSNNQIVLSVAGIFLVLFIEIMLLLHGWTIFRMWKPIIKYICLVLSVFFAIATVVMRAVY

QVIGLKYLLTTAAGKLMSQASDDLVVVSLFFANLALVISTIWFTAIFNVKLVSHMWKTKRILPTRK

GMTPMEILTVTNGLLMIVPFASSSTSNTPTLPLFNGGQDLTMTPATNKTMVETGAAGKFPNDCDNC

VNVHRDIDQRSERRLAHN*
```

B. spectabilis, *MAT1-1-9*, MAT1-1-9 protein

Gene sequence:

```
ATGAAGTCGAAGAAACCTTGTTTCTGTCATGCCTGGAAATGCGACAGAGGGCAGCTAGATGGTATG

GATATCTCGGCATCTGAAATACGTGGATGCTGACTATACTCAGAGAACGGTCAGGTGCTCCGGAAG

AGTTCTGAGCTATTTCTACTCGAACTGGAAGATACCGTGGAAGATGCCTATGCGCAAGTCTTTCAG

TTTGTCTTCCTGAGCATCCAGGAATCTGATAGGCCAATCCATAGAATTATCCCTCAGCCAGAACTC

CTGAGCTTTATTTTGGATGATATTTACCTTTCTATCGAGGAAGCTACATCTACTCTGTTAGTGGTG

GTTCCCGAGCTCCGTCGACTCCAGCGCCAGCAACATCTTTATCCATCCTTAGACTCAGCAAAGAGA

TTCCTCTGTTCTATGAAAAGCAAAATTAGCATCAACCATAAGAGAACCCTCTCAGCCGCCAAAAAG

ATGCACGAAAGTCTCTGCTGGGAAAGAATACATAGAGCGGAAGAACAGATAGTTCTAGCTATTCAA

CATCCTACTTACGACCCACATCTGCGTCCGTGGCAACAGTTTCTGATGGATGGGGCTCCCGGTAGT

GGTCTCAATGCTCGAGACATATTCTACAACAGAAAGCCAGACTCTATTACTAGCCCCTCCAAGGAT

TCGTCGTTTCGCATACTCCCCCCTGGGGCTGTATGTCTCCTAAGGGAATTACATTCAAATGCAAGA
```

```
GAGATGGAGAGAAAATGCAGAGAGGTAAATTCCTGAATTTAAACAGAAAGCAATCGTGTGATACCA
ATGTGTAGGGAAAGACCGCAATGCCAGGGATCCTTTCGTCTGGTCCGTCGCAGATACATACACATA
CAATCCAGGTACGCCAGTAGACCTCAGTATTCATTGGTATTATTGATGAATATTCGCTTACTTCTA
GAAGGTTCCAACGACCAAAGATCAGGTTCCTTCTTTCGGTCTACGACACGAAATGGTTCGAGTATG
A
```

CDS annotation:

```
join(1..43,140..724,773..819,929..991)
```

Protein:

```
MKSKKPCFCHAWKCELFLLELEDTVEDAYAQVFQFVFLSIQESDRPIHRIIPQPELLSFILDDIYL
SIEEATSTLLVVVPELRRLQRQQHLYPSLDSAKRFLCSMKSKISINHKRTLSAAKKMHESLCWERI
HRAEEQIVLAIQHPTYDPHLRPWQQFLMDGAPGSGLNARDIFYNRKPDSITSPSKDSSFRILPPGA
VCLLRELHSNAKSNRVIPMCRERPQCQVPTTKDQVPSFGLRHEMVRV*
```

B. spectabilis, *ppgA*, α-factor pheromone

Gene sequence:

```
ATGAATTTCTCTCTTGTCTTTCTTGTCGTACTGCTGGCAGTATCCGTGCAGACAATAGCCAGTCCT
CTAGATCGATGGTGCCGTCGTCCCGGCCAGCCATGTATAAAGATCAAGCGGGGAATCCAACTCTGG
AAGTCGGAACTTGGGACATCAGAACCTAGAGAGAAGGAGACTCTCGCTTTTGGACTCTTCAACAGT
CGACCGCCGGAGTTTTCTGGTCGATCTATTCATCATAGTTCAAGAGAGACGTCGAGTCTAGTTCCG
TACGCATCGCATACGAGTATAGCTTGA
```

CDS annotation:

```
(1..291)
```

Protein:

```
MNFSLVFLVVLLAVSVQTIASPLDRWCRRPGQPCIKIKRGIQLWKSELGTSEPREKETLAFGLFNS
RPPEFSGRSIHHSSRETSSLVPYASHTSIA*
```

B. spectabilis, *preA*, a-factor pheromone receptor

Gene sequence:

```
ATGGCTGACATTGAGTATAAACGTGCTCCTGAGGCCATTATTATTCCGTTTCTATCCCTCCTCTCA
ATAATACTGAGCATCACTCCGTTGATGTGGCACTTGAGAAACTCCAACTTCCCCGCTGTCTGTCTC
GTCTGTTGGTTCCTCATCAATAATCTGTTCAACATCGTGAATGCCTTCATCTGGCCTACTGACAAT
ATTGATTCCTGGTGGAGCGGCGTCGGACTCTGCGATATCGAAGTCAAATTGATGGTGGCAAGCTAC
GTTGGCATACCGGGAGCTCTGTTGTGTATCTTCCGCCATCTTGCGGAAGTTCTGGACACAGATCGT
ACAGTTCTTGTCCCTTCTGGAAGTCAGCGTCGGAAGAAGCTTGCCATTGAGATAGCATTCTGCTTA
GTCGTGCCTGTCTTGGCCATGATCACACACTACTTTGTCCAGAAAACCCGTTACATGCTCTACACG
ATTTCAGGGTGTGTGAACAACTTTGACGAAAGCTGGGTTACGTTTGTTCTATCTTACCTGTGGCCG
ACTGTGATATGCGTCATTGCAGCTTACTACTGCTGTACGGATATCCTATCGCCAATTACCAATGAT
TGCAAGATGCTAACTACTTCCTAGGTCTCGTGATGTACCGGATTATCAAGTACAACAGAGAGTTCG
GAGCTATCCTGAGTGGTTCTCGGGGGAGTAGTCTTAACAAAACGAGGTTCTTACGCCTGTTTTGCA
TGGCATCTGTGATGCTGTTGGCGATACTTCCTACCCAGGCCTACGTCCTCTATGTTGATATCCAAC
TGTCACTGCCATGGCACGATTACTCCTTTACCAGGATGCACGGTCCTGAATGGAACACAATTATCA
AGGTTGCGACCCATGGACAGAGCTTTTTCGATCGTTGGATACCAATTGGTGGAATACTTGTTCTGT
TCGTTTTCTTCGGTTTCGGAAAGGATGCGACTTCGATGTACCGTTCGATCTTCACCAAGCTTGGAG
TTGATCGCTGGCTCCCAGCAAAGTCACATTCCGATATCTACTCCGGCAAAACTACGAGCTCTTCAG
GTAGTCGAATGAGTAAACTAGGGAAAATGTTTACCTGGAGTCTCCAGTCCAACACCACAAGGTCAG
TGCCGGATTTTGCCGGATTCCTTTTGTAAGAAAACCCGTCTCCCCGCTAACTGAGAAGCAGAATTT
CACGAACTAGCTCTGTTTCCCATGATCCTTCCAGCAGTTCTTCTACTTCTACTACCGACCTGGAAA
AAGGATCCGGCATCCACTCCGTAGACAGGAACATGACCAATGAATTGTCGTTCTTCAGTAGACTCG
CCCGTCTGAGGATGTGGACGGTCTTCCGCACCGCAACAGGAAATGCCCCTGCTGGACCTCCTAATA
ATAACAGTGTTGGGGCCCCACGCGGGATGGGCGATACTCATCCAGATGCTGGTCCTTCGCCGAAAA
TAGAGTCGGGCATCAAGGTCAAGCACGTCATTCGCCAGACCAAGGTGTCCATCTGCCCACATTGCG
GTGACGAGATCGTGTCTTACGAGCCTTGA
```

CDS annotation:

```
join(1..562,619..1117,1181..1547)
```

Protein:

MADIEYKRAPEAIIIPFLSLLSIILSITPLMWHLRNSNFPAVCLVCWFLINNLFNIVNAFIWPTDN
IDSWWSGVGLCDIEVKLMVASYVGIPGALLCIFRHLAEVLDTDRTVLVPSGSQRRKKLAIEIAFCL
VVPVLAMITHYFVQKTRYMLYTISGCVNNFDESWVTFVLSYLWPTVICVIAAYYCCLVMYRIIKYN
REFGAILSGSRGSSLNKTRFLRLFCMASVMLLAILPTQAYVLYVDIQLSLPWHDYSFTRMHGPEWN
TIIKVATHGQSFFDRWIPIGGILVLFVFFGFGKDATSMYRSIFTKLGVDRWLPAKSHSDIYSGKTT
SSSGSRMSKLGKMFTWSLQSNTTSRISRTSSVSHDPSSSSSTSTTDLEKGSGIHSVDRNMTNELSF
FSRLARLRMWTVFRTATGNAPAGPPNNNSVGAPRGMGDTHPDAGPSPKIESGIKVKHVIRQTKVSI
CPHCGDEIVSYEP*

B. spectabilis, *preB*, α-factor pheromone receptor

Gene sequence:

ATGCAGAAATCTTTCGACCCTTTCACGCAGAGCGTGACCTTCCATCTGGCAGATGGATCTCCTCTC
CCAGTTTCCGTCGGAGAGCTGGACGACTTCGTTCAGTATGGTATCAAGATCTGCATCGTCTACAGC
TCTCAGCTTGGTGCGTCCCTGGCCTTGCTTGTGCTGGTTCTCCTCTTGACGAAGAGAGACAAGCGA
CTGTCCCCGGTTTTCCTGCTCAATGGCTCTGCTTTAATCTTCAATGCGTGCCGCTTGATCTGCAAT
TCCGTCTATTTCACTACCGAGTTCTCCAAGGTGTATCCTTACTTTTCAGGAGACTACTCCCGAGTT
CCAGCCAGTGCTTATGCGAACTCCATTCTGGCTGTTGTCTTCCAGGTGCTGCTGCTGGTTGCCATT
GAGTGCTCCCTCATCTTGCAGACTCAGGTTATCTGTGTCACTCTTCGTCAGCTTTACAAGAACATA
CTTCTGGGTATCTCTGTGGTTATCTCTTTGCTTGCGGTGGGATTCCGCCTGGCTCTGGCTGTTGAG
AATTCGATCGCTATCGTGCAAGCGGCCAACTTTGATGCTGTCTGGTTAGAAAGCGCCACCAATATT
ACTACCATCATTAGCATCTGCTTCTTCAGCTGCATTTTCATCGCCAAGTTGGGACATGCCATTCAG
CGACGCAAGAGGATGGGAATCAGGCGGTTCGGCGCGATGCAGGTGATTTTCATCATGAGCTGCCAG
ACACTGATTGTTCCATGTATGTTATCCGTACCGATATCACGATGGATTCTGCTAACTCATGATACA
GCCATCTTTGCGATCCTGCAGTATTGTGTCGAGGTTCCTGAAATCGACTCCTTCGTTCTCACTCTC
GTCGCGATCTCTCTCCCTCTGACCTCCCTGTGGGCGACGTCTTTGACCAGCAATGGTAGCCAAGCA
TCCAGCGGGTCAACTGAACGTAACAGACTTTGGAAGCCGATTCGTCTTGGAAGCGCAGAGAAGCTT
AAGCAGTCTTCTACGCAAGGCCCGTCCAGTTTGGCTTCTGTAGCTACTTCAAGCCCGACTGGGGCA

```
GCGCATATGAATTCCCTCTGCCCTGACCTTGAGGCTGGAACAGCGATTGGGGTGAAGCACGATGTT
ACAGTTGCTTCTAACAACATGTAG
```

CDS annotation:

```
join(1..742,794..1146)
```

Protein:

```
MQKSFDPFTQSVTFHLADGSPLPVSVGELDDFVQYGIKICIVYSSQLGASLALLVLVLLLTKRDKR
LSPVFLLNGSALIFNACRLICNSVYFTTEFSKVYPYFSGDYSRVPASAYANSILAVVFQVLLLVAI
ECSLILQTQVICVTLRQLYKNILLGISVVISLLAVGFRLALAVENSIAIVQAANFDAVWLESATNI
TTIISICFFSCIFIAKLGHAIQRRKRMGIRRFGAMQVIFIMSCQTLIVPSIFAILQYCVEVPEIDS
FVLTLVAISLPLTSLWATSLTSNGSQASSGSTERNRLWKPIRLGSAEKLKQSSTQGPSSLASVATS
SPTGAAHMNSLCPDLEAGTAIGVKHDVTVASNNM*
```

B. fulva, MAT1-1-1, MAT1-1-1 protein

Gene sequence:

```
ATGGCCGCCGTCACGACTCCCCTCCAGCAGGCGTTTAACAACTTTTTGTTGACTATGCCGCCTGAT
CGTCTGGAAGAGCTGCTCAGGTATCTCAATGAAGCCAGAGAGGGTTCAGACAATGAAACTCATGGG
GACAATACTGGGGCCGCGAATCCATTTTTTGACGCTCTTCCTCGTCCTATCTTGCCGACAACTGCA
GCCCCTCGCTCACCAGCTTTGCGTGGGAGACGTCCGCGCGAAGGGAAGCTAAGGCCTCTTAACAGC
TTTATTGCCTTCAGAAGTAAGTTAGATAATCCCAGACAGAATTGAAATACTGAGCTAAACCATGAG
CAGGCTTCTATTCGAACATCTTCCCCGACTTGACTCAAAAGGCGAAGTCCGGAATTCTCAAATTTC
TCTGGCAGAACGACCCTTTCAAGGCCAAATGGGCTGTACTCGCCAAAGCATATTCCATTGTTCGCG
ATTCTCATATTGGCGAGGTCTCTCTTGATTCGTTCTTGAATTTGAACTCCGGTTTCCTTGACATCG
TGGACCCTGCCCATTACCTTGAAGCTATGGGATGGGAGTTGACCAACCAGGGACAGAGTCAGTACA
CCATGGCTAAGGTCAGATCAGTCCCTATTGTCGAGGTCGAGGTTTCTACCATTCATTCGGTGAACG
ATATTGTCACTCGTTGTTACGAACTGGGCTACGTGTCTGGAGAAAAGCGTAACCATACGTTGAGTC
GGGGCAATGACGGCCCAGCTTCTAACATGGCATTTGTCGCTCAGCCAACATTGATTGTCAACGAGA
ACAACGCAGTCCAGATCTCTGGACCTGACAACAATGTGGTCATTGACCTCGATGCGAAAGCTGACG
```

```
CTGAAGTTTTCCCCAATAACGCTCCAGCAGGAACTCCGTCTCCGGCTCCTAGCGACTTGAGTTCGA

TGGTTCACGATAATTTCGGGACGGAGCCTGAAGTACACCATGTGATGCCTCAGTCTAAGGTTGACA

ACGGCTTCAATGCCAATGTGCAGTTTGCTGATGCGCATGACCCTGAAGTCTTCTTTCCTCTGGAAG

TGCTTCATACTCCCCTTATGCCCTACGATCCTCTCATCTCAGGGCCCATGGAGAACTACGACATTG

ACAGATTCCTGAATCTTTAA
```

CDS annotation:

```
join(1..280,334..1142)
```

Protein:

```
MAAVTTPLQQAFNNFLLTMPPDRLEELLRYLNEAREGSDNETHGDNTGAANPFFDALPRPILPTTA

APRSPALRGRRPREGKLRPLNSFIAFRSFYSNIFPDLTQKAKSGILKFLWQNDPFKAKWAVLAKAY

SIVRDSHIGEVSLDSFLNLNSGFLDIVDPAHYLEAMGWELTNQGQSQYTMAKVRSVPIVEVEVSTI

HSVNDIVTRCYELGYVSGEKRNHTLSRGNDGPASNMAFVAQPTLIVNENNAVQISGPDNNVVIDLD

AKADAEVFPNNAPAGTPSPAPSDLSSMVHDNFGTEPEVHHVMPQSKVDNGFNANVQFADAHDPEVF

FPLEVLHTPLMPYDPLISGPMENYDIDRFLNL*
```

B. fulva, *MAT1-2-1*, MAT1-2-1 protein

Gene sequence:

```
ATGGCGATGATACCTATTGCAGTGAAGGCAGCTGCCGAACCTACGGACAAAGTCACGGAGTTGATC

TGGCAGGATGCTTTGCGTCACCTCGAATCTACGAACAACGAAGTCCTGCTTCCAACGAACATCACC

GACATGATCGGCCAGGACAATCTCGACAAGATAAAGTCTCGTCTCAGGTAAGATGCCCGTTTGGAT

TACAGAGAATGGACAAGCTAATACGCAGACAGCGCCCTCATTGATGCTCCGGTTCTGGCCTTCGTC

GACAACTCGATCAATGCTCTTCGTGTCATGCGTATTCCTGCATTTGCGGGCACTGTGATCTCTCTG

TCGTCTCATACCATGATTCCCAATTTGCAGAACGAAGCCGCCGCGAACGCGGATAAGTCCGATGAA

AAATCCATCCGTGCGAGAATTTCGAGAGTTCCACGACCCCCGAATGCTTTTATCTTGTATCGCCAG

CATCATCACCCGAAAATTAGAGCTGCGTACCCTGAATTTCACAATAACGATATTTGTGAGTATACT

TAATACCCTGATACCTGGGAAGTACCGCTAATGGAGATAGCAAAAATGCTTGGAAAACAGTGGAAC

GCTGAGTCCGATGAAGTCAAAGCCCATTACAAGTTGCTCGCGGACGACATCAAGCGCAGACACGCC
```

```
GCGGAGAATCCGGGATACCAGTACGCACCTCGCAAATCATCTGAAAAGAAGCGTCGGATGGCATCG
CGCCAGTTTCAAAATGGCAGAAAGCCTAGTACAACTGCGTTGGCCGACTCCCCCGGATCCAATGCG
ACGGCCGTCTCATCCATCGTGACGACCCCAATTCCCTTTCATCCAGCTCTTGCTGAAAGAGCAAGA
AATACCGGGGCAAACGGAAACATATGTGACCTAGAACATGTGAACGTTGCCATCGGTCAAAATGAT
GTCGTGGGCGACCTGAACTTCGACTTCGACTCTGGCACCTTCGACGATTTGATCCAGCGAGCCGAA
AGCGACCAGGAAAAGGCCATGCTGTGTCAGCATATCAGTCACAAGAGTCATCAGGGAACTGGAGAC
TCTTTTAACTTCTCCGAATTTATTACTGACTGCTACTGA
```

CDS annotation:

```
join(1..179,231..517,569..1095)
```

Protein:

```
MAMIPIAVKAAAEPTDKVTELIWQDALRHLESTNNEVLLPTNITDMIGQDNLDKIKSRLSALIDAP
VLAFVDNSINALRVMRIPAFAGTVISLSSHTMIPNLQNEAAANADKSDEKSIRARISRVPRPPNAF
ILYRQHHHPKIRAAYPEFHNNDISKMLGKQWNAESDEVKAHYKLLADDIKRRHAAENPGYQYAPRK
SSEKKRRMASRQFQNGRKPSTTALADSPGSNATAVSSIVTTPIPFHPALAERARNTGANGNICDLE
HVNVAIGQNDVVGDLNFDFDSGTFDDLIQRAESDQEKAMLCQHISHKSHQGTGDSFNFSEFITDCY
*
```

B. fulva, *MAT1-2-4*, MAT1-2-4 protein

Gene sequence:

```
ATGGACTTGTACTGCCGCACTGTTGTGAGAGTTCTTGGGGCCATGGCTCAGGACGCGACTTTGCCC
TCTGAGCAAAGACACCAGGCCCTGTACGCGGCTGTCTCGTTCTTCAGTCACCGTAACCCCTTTCGG
CTTATGGCCCAGGTATCTGCTATAAACGATGGCAGATCGTTGATCCAGCCTCCCCATCAGACTACA
GCGGTGGCGGAGAATGGCGACTCCCCAGTCAACCGCTACGGAATGTTTATCAATGCCATTATTACC
GACTCTCGTGTCGTCCCCACGAATGCGGACTTCGAAGGCCACCCGATTGAACTGTTTAGTACACTG
GACCCTCGTATTGAAAATGCGTTGAATGGTGCCCAGAAATTTCAAGTTCACCGGGCCTTGCTTTCT
TTGGAGCAGGATGCCAATCAAGACCTCGTCAGATGTACCAGGAAGTACGGCTATCACTACATCTTC
AGGGCTGGATTGAGGCAATATTACATGACGTAAGTTCAAGGAACAATGAATATTTGAAGTTGCTCA
```

CGCTTTAGCAGAAAAATTGTCGCGGAGAACTTGAATTTCTGGAGAGCGGACGCGCGTGGCAATGAG

TACAGAGTCAGCGCACAGAAACTCTGCTATGAAGCAATGGAAAAGCAGGTCAAGCTGAATGCTACT

GAAAAAGAATTCTCATACAGGCGCTCGGTTGCTCGGCAGAGGACGCGCACAGATTCTGTGAGTAC

CTCTGAATTAGACATGTCGAACGCGCTAACGCTAGTAACCAGGGAGCTGGCTAGAGAGGAATCGAG

GCGCCTACATTGCAATGAAGACTTGCATTTCTCTGCTAGAGCACCTCAACTGTAACGATTCCACAC

AGGACACATAG

CDS annotation:

```
join(1..491,540..718,769..869)
```

Protein:

MDLYCRTVVRVLGAMAQDATLPSEQRHQALYAAVSFFSHRNPFRLMAQVSAINDGRSLIQPPHQTT

AVAENGDSPVNRYGMFINAIITDSRVVPTNADFEGHPIELFSTLDPRIENALNGAQKFQVHRALLS

LEQDANQDLVRCTRKYGYHYIFRAGLRQYYMTKIVAENLNFWRADARGNEYRVSAQKLCYEAMEKQ

VKLNATEKRILIQALGCSAEDAHRFWSWLERNRGAYIAMKTCISLLEHLNCNDSTQDT*

B. fulva, *ppgA,* α-factor pheromone

Gene sequence:

ATGAGGCTCTCTTCAATTCTTCTTCTGACCCTGGCTGCCGCCGCAGTGCAGGCTCTCCCAACCGGC

CTCGAACGCCCAGCCACCGAGAGCGATATCACATTGGATGCCCGTAACCATCCTCCTCGTCCTAAA

GGCTCCTACGGATACAAGAATTGGTGTGTAGACCCGGGTCAAGTTTGTGAACCGCTCAAGGTCAAA

CGCACCCCCGACGCTCTCGAGGAACGCGAGATACTTACGGAGCGCGAGACGGATCTACAGCTTTAG

CDS annotation:

```
(1..264)
```

Protein:

MRLSSILLLTLAAAAVQALPTGLERPATESDITLDARNHPPRPKGSYGYKNWCVDPGQVCEPLKVK

RTPDALEEREILTERETDLQL*

B. fulva, *preA*, a-factor pheromone receptor

<u>Gene sequence:</u>

```
ATGGGCCCTCCGATCGCTAAATTGAGGCCAAATCCGCACGCACCGCTAGAGACACATCCCTCTGCT
ATCGCTAGGGCCATACCGGCTCTGGCCTGTGTCTCTATATTCGTCACTATCCCAACTCTGATCTTA
CACCTGAAAAACCGCAACTGGCCCGTTACGTGCATCACCGCTTGGTTTATTATAGCCAACCTCTTC
TCTATGGCAAACGCCCTGATCTGGAGCACGAACGATCTGGATTCGTGGTGGAGCGGCGAAGGGCTT
TGTGATGTTGAAACCAGACTCATGGCAGCCAGCTACGTTGGTGTGCCAGGCGCCTTGGTCTGTCTG
TTTCGAAATCTGGCCGCCGTGCTAGATACGAGCCGTACTACTCTCGTCCCTAGTCAGGCACAACTC
CGGCGCAAGATTTTCCTGGAGATTTTGTTCTGCCTGGTGATTCCGGCCATGTTGATGGTTACCAGC
TACATGGTGCAGACGAATCGGTATGCGATCTTCGAGATCTCAGGATGCACCTCCCTGTGCGATCAA
AGCTGGGTGAGCATCGTGCTGCTCTATATGTGGCCCGGGATAATCTGCCTGATAGCCATCTACTAC
TGTTGTAGGTCCCCGTCCTACGAAAACCAACACCTTTGCTGCTTGACTAATGAGACTGTCCAGGCC
TAGTACTGCTCCGTGCCTACAAATACCGAGCCCAATTCGGAGATATCCTCAGGGCGTCTGGTACCA
GTTGGACTAAGTCCAGGTTCTTTCGTCTTTTCGCACTCGCCCTCTTGATCCTGCTCGCAATCACCT
CATTTCAGACATACTTCCTTTACGACAATGTACGGTGGCTCGTTCCGGGCTGGCACGGTTTTTCTT
GGCGCGGAGTTCACGGGCCTGAATGGAAGATTGTTCTCATGGTCTCCACGGATGGCAATAGTTCCA
TCGACTGCTGGGTGTTCATTATTGGCGGCTTTATTGTCTTCGTCTTCTCCGGATTTGGCCGAGATG
CCATTCAGATGTACCGCTCTGCGCTTTCTCGCCTTGGATGCGAGCGTTTCTGTGATTGTCTATCGC
GTCTATTTCGCCAGCGGGAAGGCGTACTCCCCATTTCAAACTCTTCAAAAAATAAAAATACCAGAT
CGACACGGGAACAAAAATGGCCAATATCACAGTATGTTCTATTTTTGTTTTTTTTTTGTTTTATAT
ATTTGTTTTGTGGGATCTCTGCCGGCCGTGAAGTGCTAACAACGTAACACCAGGAACTCTTCCAGA
CCGGCTCCAGATACTCTTGACCCAATACGTCAAGACTCCGGATACCGCATTCCATCAAGGCAAAAC
GAGACGAATGCGAATAATACTCCTTCCTGGCCAAGAAGAGCGTGGAAGTCCTTCCGAACACCCTTG
TCTTCTTCTCCCTCTGCATACAGTGACAACTTCAGGCCCACAGCTGCTTCCAGGCTTCATCCCACA
TTGAACAGTGTCTCTGCCAGTGCGTGGGCAGATAGAAGCCAAAAGAGCCACGGCAGTAGCACCTTG
```

```
CGTTCCTCGTCGTCTTCTGCAGAAGAGGGGATTATTCAGGTGAGGCAAATGATCAGTCAGAAGAGC

GATGTTGCGAACCGTGATGTCTAG
```

CDS annotation:

```
join(1..598,658..1180,1257..1608)
```

Protein:

```
MGPPIAKLRPNPHAPLETHPSAIARAIPALACVSIFVTIPTLILHLKNRNWPVTCITAWFIIANLF

SMANALIWSTNDLDSWWSGEGLCDVETRLMAASYVGVPGALVCLFRNLAAVLDTSRTTLVPSQAQL

RRKIFLEILFCLVIPAMLMVTSYMVQTNRYAIFEISGCTSLCDQSWVSIVLLYMWPGIICLIAIYY

CCLVLLRAYKYRAQFGDILRASGTSWTKSRFFRLFALALLILLAITSFQTYFLYDNVRWLVPGWHG

FSWRGVHGPEWKIVLMVSTDGNSSIDCWVFIIGGFIVFVFSGFGRDAIQMYRSALSRLGCERFCDC

LSRLFRQREGVLPISNSSKNKNTRSTREQKWPISQYVLFLFFFWAPDTLDPIRQDSGYRIPSRQNE

TNANNTPSWPRRAWKSFRTPLSSSPSAYSDNFRPTAASRLHPTLNSVSASAWADRSQKSHGSSTLR

SSSSSAEEGIIQVRQMISQKSDVANRDV*
```

B. fulva, *preB*, α-factor pheromone receptor

Gene sequence:

```
ATGGCATCTGAATCATTCGACCCGTTGAAGCAGAATGTCACTTTCCATTACGATGATGGGACTCCG

TTCGACGTGCCGATGGACCAATTGGACAAGTTCTCCCAGTACAACATTCGGATGTCCATCAACTAT

GGAGCCCAGCTTGGTGCCAGTCTTATCGTCCTGGTCGCTTTCGTTACTTTGACCAAGCCGGATAAA

CGCAAGTCAGCTATCTTCTTGCTGAACGTCTCTGCATTGCTCTCCAATATTGGCCGGCTGTTTTGT

CAACTCTTCTATTTCGTGGGCCCCTGGCATGAAATATATACCTATTTCGCGTATGACTATTCAAGG

ATCCCGGACAGCGCCTACGCCAACTCTGTCCTTACCGTTATCTTCAGCATCGTCCTGCTAATATTC

ATCGAGCTCTCGTTGGTCTTTCAGCTTCGGGCTCTGTGTGCGACGCTCCGGCGTCGTTACCGGCTC

CCACTTCTCATTGTGTCGTCCGGCCTGGTTCTTCTTTCAGTTGGCTTTCGGATTTGGCTCGGTGTC

CTCAATTGCTTGAATATCGTGCGATCGGTGTACTCGACGGATTTGATATGGTTGCAGAGAGCGAAC

AATATCGTTCTGACCGTCAGCATCTGCTTCTTTTCAGCGGCTTTCGTCGTCAAACTGGGCTATGCG

ATCAAGGCTCGGAAGCGACTGGGTATAACGGATCTTGGGCCGACGAAAGTTCTGTTTATTGTCGGA
```

```
TGTCAGACCATGCTTGTTCCAGGTAGGAGCATTTCTTTCTTCCTTGTTGTCTCTCTATGTCTCTCT
CCCCCCTTTCTACTCCAAGTTCAACTACAAAATGATCCAGACCAAACACATCCGCATAAATCACCA
AACTCGTTCCTAATACCAATGACAGTGCTATTCGAGACCATACAATACGGCACCACCTTCCCAGAA
ATATCCGCCAACTGCCTGACAATTATCTCGATCTCGCTTCCTCTTTCCTCGCTCTGGGCCAAGCAC
AGTCTCGACAGCAATCTCGCAAAAAGGAAAGCTGCTCTTTCCCCGCCATCGATGGGCATGGGGCAC
TCATATAACTCCAACGCTACGTGCTTTTCAGGGAAGAATCACCCGCGCCACTATGACCTGGAATCC
GGACAGGAGAAGGGCATTTCTATCGCCCGCGACATTTCAGTCCTTAGCTTCAAGAACAGTAATAAC
AGCCAGCCTAGTGGCCACCGTGTGGATATCCATCAGCCCATTTAG
```

CDS annotation:

```
join(1..748,884..1233)
```

Protein:

```
MASESFDPLKQNVTFHYDDGTPFDVPMDQLDKFSQYNIRMSINYGAQLGASLIVLVAFVTLTKPDK
RKSAIFLLNVSALLSNIGRLFCQLFYFVGPWHEIYTYFAYDYSRIPDSAYANSVLTVIFSIVLLIF
IELSLVFQLRALCATLRRRYRLPLLIVSSGLVLLSVGFRIWLGVLNCLNIVRSVYSTDLIWLQRAN
NIVLTVSICFFSAAFVVKLGYAIKARKRLGITDLGPTKVLFIVGCQTMLVPVLFETIQYGTTFPEI
SANCLTIISISLPLSSLWAKHSLDSNLAKRKAALSPPSMGMGHSYNSNATCFSGKNHPRHYDLESG
QEKGISIARDISVLSFKNSNNSQPSGHRVDIHQPI*
```

B. nivea, *MAT1-2-1*, MAT1-2-1 protein

Gene sequence:

```
ATGGCCACAGTTTCTGCGGCCGTCATAAAATCCGCTTCTGAATCGATGGACCAGGTAACCGAATTA
CTCTGGCAAGACGCTCTGCATCACCTTGGTGCTACTAACAACGAGGTTCTGCTGCCAATCAACGTC
ACCGATCTGATCGGTCAAAGCAATGTCAACAAGCTCAAGGCGCGTCTTTGGTAGGATGCCCGAACC
TCTATAGCTAGAAAGGATACTAAAGCATGAGAGTAGTGCCCTTATTCGTGCCCCTGTTGTGGCATT
TGTTGACGAATCTATCAACGCGCTTCGCATCATGCGTACTCCAGCTTTCGCCGGCTCGGCTGCTTC
TGCAGCTTGTCATGTTCTGACTCCGGCTCTCAACGTACCCATCCATGATGAAGCTGGCTTCAATGT
GGTGAAAGAAGCTGGAGATTTCGGGACCAATTCCAAAGGAGCCAAGATTCCTCGTCCTCCCAATGC
```

```
GTTTATCCTCTACCGTCAGCACCACCATCCGATCGTCAAAGCGGCGCATCCTGAGTTTCATAACAA
TGATATCTGTAAGGTGTCACATCACCTATAATAGTTGAAGTAATGCTAAAATTTATAGCTGTCGTT
TTAGGAACAGTGGAAACAAGAGCCTCAGCATGTCAAAGCTCACTTCCAAGCCCTTGCAGAGGAG
ATAAAGCGCAAACATGCTGAAGCCTATCCAAACTACCAATACAGTCCTCGCAAGTCTACCGAGAAA
AGGCGTCGGACTCTTACTCCGCGCCAGCCTCTCTGCAGGCCGAGTGATCTCATGTCGATACCACAA
GACACGTCGCCAATTGAGTCGACCACCGCAGCTTCTGTCAAACAGTCCATTAGAACGGCCATGCCT
ATTCAGATTGATTCGATTCCTTTCCAGGATCTTGACTCTCACGTGGAGAGCCAGGTGTCCGATTTT
TTTTTCGGGGAAATGATTGACTACGAAACATCGAGTTTGACGAAGAAGAATTCGAATCTATGATT
CAGCAGGCTGAGAGCGCACAGGACAGGGCAGCACTGTTTGAAGAAATCACCATCGTCCCAGCTGCC
ACTCCTGACCCTAACGACTGGATGAACTTCATTATGTAA
```

CDS annotation:

```
join(1..182,235..536,587..1095)
```

Protein:

```
MATVSAAVIKSASESMDQVTELLWQDALHHLGATNNEVLLPINVTDLIGQSNVNKLKARLCALIRA
PVVAFVDESINALRIMRTPAFAGSAASAACHVLTPALNVPIHDEAGFNVVKEAGDFGTNSKGAKIP
RPPNAFILYRQHHHPIVKAAHPEFHNNDISVVLGEQWKQEPQHVKAHFQALAEEIKRKHAEAYPNY
QYSPRKSTEKRRRTLTPRQPLCRPSDLMSIPQDTSPIESTTAASVKQSIRTAMPIQIDSIPFQDLD
SHVESQVSDFFFGEMIDYENIEFDEEEFESMIQQAESAQDRAALFEEITIVPAATPDPNDWMNFIM
*
```


B. nivea, *MAT1-2-4*, MAT1-2-4 protein

Gene sequence:

```
ATGTACATATCAGCCTTTTTTCAATGTTCATCACAACGCCACTCGTCTCGTTGCCCAAGTTTCAGAG
TTCACAGCTGGAGATATGGATATCGCTCATTGCCATGGACTAGGAAATACAACCAGAATGAAGGAT
AAGGCAGTAGACTACCATATTGAACTCCTTCAGTCGATTCTTACAGATTTCGGGGGTTGTTCCTAAT
ATCGCAGACTTTGAGGGCCATGCTATAGAGTTGATTAGCACTCTTCCTGCAGGAATTGAAAATTTA
CTTGGCCCTGAGAAAAGGCTCGATTTTCATAGGGATCTACTTCATCTCGAAAAAGAAGCAAATGAA
```

```
AATCTTAACGAATATACGAACAAGTATGGTTATCACTATATCTTCAGAACCGGACTTCGAAATTAC

TGCACGACGTTGTTTTTACAAAACTTGATCCATAATCAACATAAAACAGACGCGAAATAGGAGGGG

TGTCGTCAAGCATATCAATTTCTGGAGGCCGGACCCTCGCGGTCCAGGATATCGAGAGTTCGCTCA

GAGACTATACTATGAAGCTATCGAAAAACACCCGGCACTCAATCCTGCAGAAAAGCGTGTCATCAT

TATAGCAATAAGGTGTTTCCCAGAGGACGCTTTGAGATTTTGTGGGTATGATCCTCCACTAGTTCT

CAGTTTGTCTAACACGTGTATCAGGGAACTGGATCACTAAACACCGTATGGCGTATTATGCGATGA

GAACATGTACCGTGTTACTTGAAGAGCAGCAATGCTTGAATTGTACGTACAGAGAGTTTACAATGC

TGATTTTACTGAGGAACCCAGAACTATTTCGATGA
```

CDS annotation:

```
join(1..404,451..635,685..827)
```

Protein:

```
MYISAFFNVHHNATRLVAQVSEFTAGDMDIAHCHGLGNTTRMKDKAVDYHIELLQSILTDFGVVPN

IADFEGHAIELISTLPAGIENLLGPEKRLDFHRDLLHLEKEANENLNEYTNKYGYHYIFRTGLRNY

CTTNRRGVVKHINFWRPDPRGPGYREFAQRLYYEAIEKHPALNPAEKRVIIIAIRCFPEDALRFWN

WITKHRMAYYAMRTCTVLLEEQQCLNCTYREFTMLILLRNPELFR*
```

B. nivea, *ppgA*, α-factor pheromone

Gene sequence:

```
ATGAAGCTCTGCCTTATCCATGTTCTCATGCTGCTGGGTGTATCTCTGAAGACGACAGCAAGCCCC

CTAGTCAATCTAAATCATTGGTGCCGTCGTCCCGGGCAGCCCTGTGCAGTGAGAAAACGTGCCGTT

GAACATTGGCAGTCGGAGTCCGAGTCGTCTAAGCTTAATGACCGGCAGCCTTTTACTTTCGGACTT

TTCAACAGTCAATCTCCAAGCTCTTCGAGTAGACCAATTGAAGGTATCTCCAGGGAAGCATCGATA

TCGTTGCGTAAGCATATGGACCATCAATTGACGGGCATGTGGACCGACTGCCAGGCATAG
```

CDS annotation:

```
join(1..324)
```

Protein:

```
MKLCLIHVLMLLGVSLKTTASPLVNLNHWCRRPGQPCAVRKRAVEHWQSESESSKLNDRQPFTFGL

FNSQSPSSSSRPIEGISREASISLRKHMDHQLTGMWTDCQA*
```

B. nivea, *preA*, a-factor pheromone receptor

<u>Gene sequence:</u>

```
ATGGTTTCCAGCAATTACGAGCGTGCTCCTGCGGTCATTATCATCCCAGTCCTATCGCTCCTGTCA

ATCATACTGAGCATGACGCCCTTAGTCTGGCACTTGAGGAACTCCAACTTTCCTGCTGTGTGTCGG

TTGGTTTCTGGTCAATAACGTCTTCAACATTGTGAACGCCTTCATCTGGCCTACGGATAATATTGA

CTCATGGTGGAGCGGTGTTGCACTCTGCGATATCGAAGTCAAGCTGATGGTAGCCAGTTATGTTGG

CATGCCTGGAGCCCTGCTATGCATCTTTTGCCACCTTGCGCAGGTTCTGGATACGGACCGTACAGT

TCTCGTCACTAGTCACAGTCAGCGTCGGCAGGAACTTGTTATCGAGATAACTTTTTGCTTGGCCGT

GCCTGCCATAACTATGATCACCCACTACTGTGTTCAGAAGACTCGTTATATGCCCTACACGATTTC

GGGCGGTGTAAACAACTTCGACGAAAGCTGGGTCACTCTTATGCTTTCTTACCTGTGGCCAAGAGT

GATTTGCAATAATTGCGGCTTACTATTGCTGTATGACTTTCCTATCCTTGGTATGACGACCGAAAT

ACTAACTACTACCAGGCCTCGTGATATACCGTATCGTCAAGTACAACAGAGAGTTCGGAGCCATCC

TTAGTAGTTCCAGCGGAAGCAATCTGAACAGGACAAGGTTCTTACGACTGTTCTGCATGGTGTCTG

TGATGCTACTGGCAATTCTTCCCACACAGGTCTACGTCTTTATGTTAATGTCCAGCTATCACTGCC

ATGGCACGAATACTCCTTTGCCAAGATCCACGGTCCCGACTGGAATATCATCATCAAGGTTGCAAC

TCATGGCCAGATTTTCTTTGACCGCTGGATACCAATTGGTGGGGCGCTTCTTCTATTTGTCTTCTT

CGGCTTTGGAAAAGATGCTACTTCCATGTACCGTTCGATCTTCGTCAAGTTGGGGATGGAGCATTT

CCTCCCATCTAAGTCAAATTCCGATATCTTCTCTGAAGAGTATGAGCTCCTCCGGGAACAAACT

GGGTTCTCTTAGCATTCGTGCTAAACGAGTGTTTACCTGGGGACAATCCACCGTGTCAAGGTCAGT

TTGTGCCTTTCCTGTGAGAAATCATTTGTCCTCTAACTTGAATGCAGCAGCATGAGAACGGGATCT

GTCTCTACTGATCCTTCCACTCCCCCTGCTACCGATTTGGAGAAAGGACCCCGCATCCACTTCACA

GAAAGGAACATGACTAAAGCTCTACGATCCAGAATGGCTCCTCCGCGAATCTGGAGTCTTTTCCAT

GCCCAACGAGGAAGCAGCTCTACTGTGCCTGAAAATAACAATCTTGCTTCTACGGAAAACACTGTT

TCTACCAGTGCCTGTGCAGGTCATAGTCCGTGTGAGATGGGTGATGCCGTCCCATCACCGAAAACA
```

```
GAGGGGGCTATCAAAGTCAAGCATGTCATTCGCCAGGCCAAGGTCTCCATCTGCCTATACTGCGGC
TATGAGATCGTGTCTTACGAACCATGA
```

CDS annotation:

```
join(1..558,610..1116,1167..1545)
```

Protein (Truncated):

```
MVSSNYERAPAVIIIPVLSLLSIILSMTPLVWHLRNSNFPAVCRLVSGQ*
```

Note: The red stop codon is the in-frame stop codon that disrupts this gene and produces a truncated protein.

B. nivea, preB, α-factor pheromone receptor

Gene sequence:

```
ATGGAGAGATCTTTCGATCCCTTCACACAGAATGTAACCTTCTACTTAGCCAACGAATCTCCTCTT
GAGGTCTCCGTCGGAGACTTGGACGATTTCATCCAATACGGCATCAAGATCTCCACCGTCTACAGC
TCTCAGCTTGGTGCTTCGCTGGCGCTGCTCGTCCTTCTACTCCTCCTGACCAAGCGTGACAAGCGA
GAGTCTCCAGTCTTCTTACTCAACAGCTCAGCTCTGATTCTCAATGCTTGCCATTTGATCTGCAGC
TGCGTCGATTTCACTACTGAATTCTCGAAGGTCTATCCTTACTTCTCCGGAGACTACTCTCGGGTC
CCGGCCAGCGCCTACGCCAATTCCATTCTCGCGGTTGTCTTCCAAGTGCTGCTGCTGATCGTCATT
GAATGCTCCCTTATCTTGCAGACTCAAGCTATCTACGTCACGATTCGCGACGTCCAGAAGCATATA
CTGCTGGGTGTCTCCGTTCTGATCGCTCTCCTTGCAATAGGGTTCCGCCTGGCCCTGGCTGTTGAA
AACTCCATTGCTATTGTCCAGGCGGCGAGTTTCAACGCGGTCTGGTTGGAAAGCGCTACCAACATC
ACCACAATTATCAGCATCTGCTTCTTCTGCGTCCTTTTTATTTTCAAGCTGGGACATGCCATCCGA
CGGCGCAAGATGCTGGGGATCAAACAGTTTGGAGGGATGCAGGTGATCTTCATCATGAGCTGCCAG
ACCTTGATTGTCCCATGTATGTCTCTCTCAGCTTAGGTTGTATTGAAATGAACTGACTTGTCATAC
AGCTATCTTCGCGATCCTGCAGTACTGCGTTGACGTCCCTGAAATGGAGTCCTTCGTTCTTACTTT
TGTCGCGATTTCTCTCCCACTTACCTCTCTGTGGGCGACATCTTTCACCACCAACCGTAGCGAGCC
AGGAAGCGAATCTAGTAGACGCTATAGGTTGTGGAAGCCGCTCTCCAACGGAAGCTCAGAGAAGAA
CGGACGGTCTTCTACGACGGAGTCTTCTGCATCGGGTGCCCTCAATACGACCTCTGTGGGTAGTTC
```

```
AAGCGCGACTATTGCGCAGATGAATACTCTCTACCCTGACCTCGAGGCGGCACTGAGATCGGTATC

AAGCAGGATGTGA
```

CDS annotation:

```
join(1..742,795..1135)
```

Protein:

```
MERSFDPFTQNVTFYLANESPLEVSVGDLDDFIQYGIKISTVYSSQLGASLALLVLLLLLTKRDKR

ESPVFLLNSSALILNACHLICSCVDFTTEFSKVYPYFSGDYSRVPASAYANSILAVVFQVLLLIVI

ECSLILQTQAIYVTIRDVQKHILLGVSVLIALLAIGFRLALAVENSIAIVQAASFNAVWLESATNI

TTIISICFFCVLFIFKLGHAIRRRKMLGIKQFGGMQVIFIMSCQTLIVPSIFAILQYCVDVPEMES

FVLTFVAISLPLTSLWATSFTTNRSEPGSESSRRYRLWKPLSNGSSEKNGRSSTTESSASGALNTT

SVGSSSATIAQMNTLYPDLEAALRSVSSRM*
```


T. aurantiacus, *MAT1-1-1*, MAT1-1-1 protein

Gene sequence:

```
ATGGCTGCGAACGTTTCTCCTCTTCAGCGTGCTTTCAACAATTTTCTTACGACAATCCCACAGAAG

GAGCTAGAAGAGCTGCTCGAGTACCTGCATAGTCTTAATGCTGGAGGAAACAGTCCAATTCCTCCG

CATGCTGAAGGTGCTCCTGCGGCTATAGCCTCTGTTGAAAATAATAACCGTCGTCCAACCTCAGCC

CTCAATGCTGGTATTCGACCGATGTCTTCGCGTGGAAAGCGCTTCCGCGAGGGGAAGCTTAGGCCC

CTTAACAGCTTCATTGCTTTCAGAAGTTTGTGACATCCCTACTAAACAATGAGAAAAAAAGAACT

GACTAATCCCACAGACTTCTATTCGGTTATCTTTCCGGATCTCACTCAGAAGGCAAGGTCAGGGAT

TCTCAAGTTTCTCTGGCAGAATGATCCATTTAAGGCCAAATGGGCCATTCTCGCGAAAGCTTACTC

CATCATTCGCGACAATCATGTGGGCGAAGTTTCTCTTGATTCCTTCCTTAACCTGAATGGAAACTT

CATCGGAATTCTGGAGCCAAATCGCTATCTTGCAGCTATGGGCTGGGAGCTCATTGTCGATGACCA

GCACCAGTACACCATGGTGAGGGTCAATCCGACTTATCCAACCGAGACTGAGGCCGCGACCAATTA

CTCGGTCAATGATATTGTTAATCGCTGCTACGAGACAGGTTATGTTTCTGGTGATCCTCGTCCATG

CAATATCAACTACCAGGGAACTAGCCCTCTGATGGCGTTCGCTGCTCAGCCAACTGTTGTCAATGG

TAGCAGTGATATCCAGGCTACTGCACCCAAAGGAATCGTGGGTGGCAACGAAAACACAAGCATGGA
```

```
CTACAATGAAAATTTCGATGCGGATGCCATCAATCACGCTCACTACATTCCTAGCGAGGTGGCACG

CGCAGGTACTGATGAGGCTCGTCCTGCAATTCCTGAATTACTCAACGTCCGCAGTGGCCCGCATAC

GCAGGAGGAGTTCATTGCTGAACTTGACGACGCAATTCATGAAATGCGAATCCCAAATCCAAAGGA

CGACGAGTTGTTCGTGCCATTCAATCCAGACATCCAGTCTCCAATTTTGCGCTATGATCCAATGGC

GCAGGAACCATTAGATGCCTTCGATATCAATGAATTTGTCAATTTTTGA
```

CDS annotation:

```
join(1..289,345..1171)
```

Protein:

```
MAANVSPLQRAFNNFLTTIPQKELEELLEYLHSLNAGGNSPIPPHAEGAPAAIASVENNNRRPTSA

LNAGIRPMSSRGKRFREGKLRPLNSFIAFRNFYSVIFPDLTQKARSGILKFLWQNDPFKAKWAILA

KAYSIIRDNHVGEVSLDSFLNLNGNFIGILEPNRYLAAMGWELIVDDQHQYTMVRVNPTYPTETEA

ATNYSVNDIVNRCYETGYVSGDPRPCNINYQGTSPLMAFAAQPTVVNGSSDIQATAPKGIVGGNEN

TSMDYNENFDADAINHAHYIPSEVARAGTDEARPAIPELLNVRSGPHTQEEFIAELDDAIHEMRIP

NPKDDELFVPFNPDIQSPILRYDPMAQEPLDAFDINEFVNF*
```

T. aurantiacus, *MAT1-1-9*, MAT1-1-9 protein

Gene sequence:

```
ATGGTAGAGGGCAGGTTAATGGTTAGTGTTTCATTTCATATCGACGAGTAAGTTACAGTAGATTAA

CGGATCACTTAGGAGAAGGCCTCGTGCTTGAGAAATGTGACGCTCTGTTTCTTATGGAGCTGGATG

AAGCGGTGAAAATACCTACGAACAAGATTTTCAGTTCATTTTGTCGATATCGAAGAGTCCGACGGT

CCTTATGGCGAAATTCATCCTGACCCCGGACAGTTCGAGCTCATTCTCGAGGACATTTATCTTGCT

ATCGAAGCTGCTTACGTTCTCCTATCGGTCGCTCCTGAAATCCGTCGTATTCAACGATATCAAGAT

CTGTACCCAACGGAAGAACTGTTGCAAAAGTTTCTCTGTTCCATAAAGAACAGCCTTACAGTGAAC

CACAGACGCACTCTGGCTGCCGCAAAGAAGATGCGGGATGAGTCTTACTGGGAAAAAATCCATAGA

GCTGAGCGTCTGGTTGTTCTGGCAATTCAACGTCCAACGTTCTATCCCCACCTCAATCAGTGGCAG

CAGTTCCTTATCAGTGGTGCACCAGGAAGTGGCATTGACGTGCGCAATATGTTCTACAGCAGGTTA

CCAGAAGAAGCTCGGACGGCATCTGAAGCTTCACCAGATTCTGCGTTCCAGACCCTCCCAGTTGGC
```

```
GCAATGGCCCTTTTGGGAGAGTGGCGGGCTAGAGCAAGGGGCCTCGAGAGAAAATGCAGAGAAATG

AGCCACATATATTGGACAAGTAAATTTGAGGATGCAAGTTCTCACTCTATCCTCCTTCAGCCAGCT

GTTCCGGAGAACATGCAGCAACATGATTCAACAAATACAGTAAGCTTCTAAACAATCTATTGAACA

GCCATAATTTTTGAACACATCCTTAGAACAGTACATCTTCTCGGATTTCTCTTCCATCCTTTCGGG

TCAAACATAAATACTCCCGGTATGATTACCCTCTAAGAATCACATAGTCGCTTCCACTAACAGTT

ATTTTAGTTTACTTATGATCCTAATATCTGGTTCCCTGTATGTCGAGTTCCTCTTTACTCATTCCT

TCATAAACTTGAGTTTTTGATTTTCAACATTTTAACTGACATCCCACAGCGTCAGCCAAAGATTCC

AAATACCTCGTCTACTGTAAGGCTTGAAGATCTTTCTCTGATGATCGGTATGCTTCAAGCTCTACG

TTCCAATTTAAAAGTCTTCACTAACTTATATAACCTGCAGACCAAATAGAGGAATTGCAACCAAGG

CGTGCGTATGCGATGAGCCACGCATATGGTGTGATCATTCGAACGGAGTGTTCCTCTGATGGTAGG

ATCATTCATCATGAAGTTCCTCTTCTGTGA
```

CDS annotation:

```
join(1..47,159..831,885..944,998..1027,1106..1169,1229..1350)
(1..289,345..1171)
```

Protein (Truncated):

```
MVEGRLMVSVSFHIDE*
```

Note: The red stop codon is the in-frame stop codon that disrupts this gene and produces a truncated protein.

T. aurantiacus, ppgA, α-factor pheromone

Gene sequence:

```
ATGAAACTTCATTCCGTCCTCGCTGCCATCCTGGCAATGTCGTGGAGGCAGTAG
```

CDS annotation:

```
(1..54)
```

Protein:

```
MKLHSVLAAILAMSWRQ*
```

T. aurantiacus, preA, a-factor pheromone receptor

Gene sequence:

ATGGCTGCCGGAACTTATGCCCTTTCTGCTCAAGCTGTGGTCGTACCGGCCCTGTCTGTCTTTTCG

ATCATTCTGAGCATCCCTCCCTTGGTCTGGCACTCGAAGAATCACAACTTTCCTGCTTCATGCCTT

GTCATCTGGTTCCTTATCAACAACATTTTCAACATCGTGAACGCCGCTATCTGGCCGACAGATGAT

GTCGATTCGTGGTGGAGCGGCGTTGGCCTCTGCGATATCGAAGTTAAGCTGATGATAGCCAGCTAT

GTGGGCGTACCAGGAGCCCTGATGTGTATCTTCCGTACCCTCGCCCATCTCCTGGACACGGACCGT

GTCGTTCTGGTCCCTAGCCGCGAGCAGTCCCGGCGGCGCCTTGCCTTCGAAGTATTCTTCTGCTTG

GTTGTGCCAGCCATCTCCATGATCACGCACTACGTCGTCCAGAAGAATCGATTCATGCTCTACACG

ATTGCCGGTTGCGTCAACTCCTTCGACGAAAGCTGGGTGACGCTCGTGCTTTCTTTCATCTGGCCC

ACAGTTATCTGCCTGATTGCCGGGTACTACTGCTGTATGGCCTTCTTTTTCCCCTTTGATCTGAGT

ATCCAATTCTGACGAGCCGTTTACAGGTCTGGTGCTATTCCGCCTCTGGAGGTACAAGAATCAATT

CGGGACGATCTTCAGGTCTTGTGACAGCAGCCTAAGCAAATCACGCTTCCTACGACTCTTTTTCCT

CGCATTCATGACGCTTCTCGTTATCCTGCCTGCGCAGGCTTACGTTCTCTACACCAACGTGGAATT

CTCGCTTCCATGGCACTCGTACTCGTGGAGCAAGATACACGGACCGAGCTGGAACATGATAACCAA

GATCGCGACGCACGGCCAGATCTTCTTCGACCGTTGGATTCCCATTGCCGGCGGCATCCTGCTATT

CTTCTTCTTCGGATTCGGAACGGACGCAACCATGATGTACCGCTCGTTCCTCCTGAGGCTGGGACT

GGGCCGCTGCTTTCCCGCTCTGTCACACCCTCACATCTGCAGTAGCCGTAGGACGAGTACAAGTTC

TGTGAGCCCAATAGTATCATTCAGCAGCCGCGCGAAGCTGCTATTCAGTCGGACTCAATCTATCGT

GTCGAGGTGAGACTCCCTCTCTTATCATCCTTCTCGGATCCAATTAGGTATTTACGTAGGAGTGGC

AGGACAAACTCCGTATCCACTGAACCCTTCGCGCCCTCTTCTTCTTCTTTTCCTTCTAAGGATCTC

GAGAAAGGATTCCCTTCCATTTCCAATACCACTACTACATTTCCACGCCGTCACCAACTCTGGGGC

CTCTTCACCCCCAACCGCGAGCGTCGCCGTAGCCGTCACCGTCACCGTGACATTCCCATCCCTCGC

CTCCCAGGCTCAGAGAATACGGTTTCCACCACGGCTTGGGCAGGCAAGGGCGAGCGGCGCTCGAGC

GAGGCCGATACCGACACCATCCCGTCTCCTTCCAAGAATGGCATCATCAAGGTGAAGCAGATGATT

CGGCAGACAAGCGAGAAGAGGTGA

CDS annotation:

```
join(1..562,621..1128,1182..1542)
```

Protein:

```
MAAGTYALSAQAVVVPALSVFSIILSIPPLVWHSKNHNFPASCLVIWFLINNIFNIVNAAIWPTDD
VDSWWSGVGLCDIEVKLMIASYVGVPGALMCIFRTLAHLLDTDRVVLVPSREQSRRRLAFEVFFCL
VVPAISMITHYVVQKNRFMLYTIAGCVNSFDESWVTLVLSFIWPTVICLIAGYYCCLVLFRLWRYK
NQFGTIFRSCDSSLSKSRFLRLFFLAFMTLLVILPAQAYVLYTNVEFSLPWHSYSWSKIHGPSWNM
ITKIATHGQIFFDRWIPIAGGILLFFFFGFGTDATMMYRSFLLRLGLGRCFPALSHPHICSSRRTS
TSSVSPIVSFSSRAKLLFSRTQSIVSRSGRTNSVSTEPFAPSSSSFPSKDLEKGFPSISNTTTTFP
RRHQLWGLFTPNRERRRSRHRHRDIPIPRLPGSENTVSTTAWAGKGERRSSEADTDTIPSPSKNGI
IKVKQMIRQTSEKR*
```

T. aurantiacus, preB, α-factor pheromone receptor

Gene sequence:

```
ATGGATATCAGTGACCCATATACTCAGAGTTTTACTCTCATAGGCGCCGATGGCACTACTTTCAAT
GCGACTTTGTACGAGATAGACGACTATCTCCAGTATAGCATCAAGATCTGCATCAACTACGGCTCC
CAGATCGGTGCTTCCATCGTACTCTTTATCGTCCTGCTGCTGTTGACAAGGCTTGACAAGCGTCAT
TCTCCCGTGTTCTTCCCCAACAGCTGCGCTCTCTTTTTCAACATCTGCCGGCTAATCTGCATGGTC
ACATATTTTACCACTAAGATTTTCAGGGCCTATCCCCTCCTCTCGGGCGACCTCTCGGAAGTCCCA
CTCAGCGCCTACGTCAACTCCATCCTGGGTGTTATCTTGACATTCTTCCTGGTGGTCTGCTTTGAT
CCTGCAGACACAAGTCGTCTGTACGACTCTCCGCAATGTTTACAAACGCATCATCTTCGTGATGTC
AATCCTCATAGCTCTTGTTCCAATCGGATTCCGTCTGGCGTTTGTAGTCGAAAACTCAAAAACCAT
GGCGTCAGAGGGTTCTACTCGATATGCCTGGCTGCTAAATGCCACAAACATCCTCATAACCATCAG
CATCTGCTTCTTCAGCGCGATCTTCGCAATCAAGCTGGGTTACACCATTTGGCAGCGCAGGCAGCT
CGGACTGAAGCAGTTTGGTCCAATGCAGGTAATCTTCATTATGGGCTGTCAGACTATGATCGTTCC
TGGTAA
```

CDS annotation:

```
(1..732)
```

Protein:

```
MDISDPYTQSFTLIGADGTTFNATLYEIDDYLQYSIKICINYGSQIGASIVLFIVLLLLTRLDKRH
SPVFFPNSCALFFNICRLICMVTYFTTKIFRAYPLLSGDLSEVPLSAYVNSILGVILTFFLVVCFD
PADTSRLYDSPQCLQTHHLRDVNPHSSCSNRIPSGVCSRKLKNHGVRGFYSICLAAKCHKHPHNHQ
HLLLQRDLRNQAGLHHLAAQAARTEAVWSNAGNLHYGLSDYDRSW*
```

T. crustaceus, *MAT1-2-1*, MAT1-2-1 protein

Gene sequence:

```
ATGGCGACTATCTCTTCCATTGCGATAAAATCAGCTGCCGAATCCATGGACCAAGTGACAGAATTG
CTCTGGCAGGACGCTATGCGTCAACTCGGCTCTACCAACAACGAGATCCTCTTACCGACCAACGTG
ATTGACTTGATCGGCCAGGCCAACGTCGACAAGATCAAAGCCCGTGTTTGGTAATGTTTTTGTTCT
ATTTTGATGACTCTAGCTCATTAAATGCAGTGCTCTAATTCGTGCTCCTGTTGTGGCATTCATTGA
CGAGTCGATCAATGCTCTCCGCATCATGCGTAGACCGGCGTTCGCAGGCTCAGCCAGTTCTGCAGC
GTCCCACGTCGTCGTTCCGCTGTATAGTGTCGCCGGCCATAAAGACATGGCTTCTGAGAATATGAC
GGCTACACCTAAAATTGGAACTCCTGCAAAGGCTGCAAAAATTCCTCGTCCTCCGAATGCCTTCAT
TCTCTACCGTCAACACCATCATCCCATGGTCAAGGAGGCGCATCCCGACTTTCATAACAATGACAT
CTGTATGTATTTCTGTCTATCTTCCCTCCTTATTTAACAAGGAATTACTGACGAACAATTAGCAAT
CATGCTCGGAAAGCAGTGGAAAGAGGAGTCAGCCGATGTCAAAGCTCACTTCAAAGCTCTTGCAGA
TGAAATCAAGCGCAAACATGCTGCAGATTATCCAGACTATCAATATGCCCCTCGAAAGCCGTCTGA
AAAGAAACGCCGGTCGACTTCTCGCAATGCTTCAACTAGAGAGCAGCGCAATCTACCCCCCTTGCC
AGTGGCATCCTCATCTGTCCCATCCGCCATTTCGGCGCCAGCTGAGCAGTCTGCTCTTACGATAGG
GGATATCGATAATATTGATTCGATGAATGACATGTTATCCTTCAACGCTGCTCTTGGCGTTAGCAC
TAATTATGAAAATGCTGAGTTTGACAATGAAGAGTTCGATGCTATGATCCAGCGCGTCCAGGACGA
CCATGATCGAGTGGCATTCCATGAAAACTTCAACCTCACATCCAGCGCCCCGTTGGATACGCTGCA
CTTCTCGGACTTTATCGTTGACTACTATTGA
```

CDS annotation:

```
join(1..182,229..530,591..1087)
```

Protein:

MATISSIAIKSAAESMDQVTELLWQDAMRQLGSTNNEILLPTNVIDLIGQANVDKIKARVCALIRA

PVVAFIDESINALRIMRRPAFAGSASSAASHVVVPLYSVAGHKDMASENMTATPKIGTPAKAAKIP

RPPNAFILYRQHHHPMVKEAHPDFHNNDISIMLGKQWKEESADVKAHFKALADEIKRKHAADYPDY

QYAPRKPSEKKRRSTSRNASTREQRNLPPLPVASSSVPSAISAPAEQSALTIGDIDNIDSMNDMLS

FNAALGVSTNYENAEFDNEEFDAMIQRVQDDHDRVAFHENFNLTSSAPLDTLHFSDFIVDYY*

T. crustaceus, *MAT1-2-4*, MAT1-2-4 protein

Gene sequence:

ATGGACCCATACGGCGCTATCCTCCTACAGTGGCTTCAAGAAATGGCTGCGGATGCAAACCTTGGG

CCGGCCAAAAGAAGTAAGGCATCGTATATCGCAACCTCATTCCATACTCATCGCAATACCATTCGC

CTTTATGGCCCAAGTTTTAGCTCTTTTTAG

CDS annotation:

```
(1..162)
```

Protein:

MDPYGAILLQWLQEMAADANLGPAKRSKASYIATSFHTHRNTIRLYGPSFSSF*

T. crustaceus, *ppgA*, α-factor pheromone

Gene sequence:

ATGAAGCTTCTTTCCATATTCGCTGCCATTCTCGCAGTCATCGTTGTGGAGGCAGTTGCGATTCCT

GGCGTCCATAACGGCGGTAATTGGTGCTCTTTAAACGGCCAGGGCTGCAAGCTCAAACGTACTGCA

GTCCCTGTCGAGGATTCCCTCGCCAAACCTGAACCCAATGCCGAGGCCGCCGACAAACACGCTGCC

TTCTGA

CDS annotation:

(1..204)

<u>Protein:</u>

MKLLSIFAAILAVIVVEAVAIPGVHNGGNWCSLNGQGCKLKRTAVPVEDSLAKPEPNAEAADKHAA

F*

T. crustaceus, *preA*, a-factor pheromone receptor

<u>Gene sequence:</u>

ATGGTCTCTGAAACTTACCCCCTTTCTGCTCAAGCAGTGGTCATACCGGTTCTGTCTTTCTTTTCG

ATCATTCTGAGTATCCCACCCTTAGTCTGGCACTCGAAGAATCGCAACTTTCCTGCATCATGCCTT

ATCGTCTGGTTCCTTATCGACAACGTCTTCAACATAGTGAACGCTGCTATCTGGCCGACTGACAAT

ATCGATTCGTGGTGGAGCGGCGTTGGCCTCTGCGACGTCGAAGTTAAGCTGATGGTAGCCAGCTAT

GTGGGCGTACCTGGAGCCCTGATGTGCATCTTCCGTACCCTTGCCCACCTCCTGGACACGGATAGC

GTCACTCTGGTCCCAAGTCGCGGACAGTGCCGAAGGAGACTTGCTTTCGAAATACTTTTCTGCTTA

GTCGTACCGGCTATCTCCATGGTGACCCACTACGTCGTCCAGAAGAATCGATTCATGCTCTACACG

ATTGCTGGGTGCATCAACTCTTTCGACGAAAGCTGGGTGACGCTCGTGCTCTCTTTCATCTGGCCC

ACGGTGATCTGCCTGACTGCCGCGTACTACTGCTGTATGGTTTTCTTTCCTTTCATTTGGATCTCC

AAGACTGACGAGCCGTTATAGGCATGGTGCTATTTCGCATCTGCAAATACAAGAATCAATTTGGGA

CGATTCTCAAGTCTTCTGATAGCAGCTTGAGCAAATCAAGATTCCTACGACTTTTCTTCCTTGCAT

TTCTGACGCTTCTCGTTATTCTCCCCGCACAGGCTTGCGTTCTCTACACCAACGTGGCACTCTCAC

TTCCATGGCACCGATACTCCTGGGGCAATATACACGGGCCGAGCTGGTATACGATAACCAAGGTGG

CTACGCACGGCCAGGCCTTCTTCGACCGTTGGATTCCTATTGCCGGTGGTATTCTGCTATTCTTCT

TCTTCGGTTTTGGAACAGACGCGACCATGATGTACCGCTCGTTCCTCTTGAGACTGGGACTCGGAC

GTTGTTTTCCCGATCTGTCGCGCCCTCGTAACTGCAGCGCTCGCAGGATGAGTTCAATTTCTGTAG

GCCCAGTAGAATCATTCAGCAGTCGCGCGAAGACGCTATTCAGTCGAACGCAGGCTATGGTTTCAA

GGTAAGGTGTTCTAACTGCACTGTGCAACTCTCCGATCATCCTTAGACCGAATTATTTACAAGCGT

ACTAGGACCAATTCTATCGCCACGGACCCCTTCACGCCCTCTTCTTCCAGCGATATTGAGAAAGGA

TTCCCCATTTCAAACGGTTCCATAAACACCATCCCACATCGTCACCAACTTTGGGGTCTATTCACT

```
CCTAACGGCCAGCGTAACCGTAACATGTCCGTTCCTCATCTTCCGGGCTCAGAGAATACGGTTTCC

ACCACAGCTTGGGCTGGGAAAGACCAGCGCCGCTCGAGCGAAACCGATGCAGAGACCATCCCGTCT

TCTGTCAAAAATGACATTATAAAGGTGAAGCAGATGATCCGGCAGACGAGCGAGAAGAGGTGA
```

CDS annotation:

```
join(1..562,616..1123,1194..1515)
```

Protein:

```
MVSETYPLSAQAVVIPVLSFFSIILSIPPLVWHSKNRNFPASCLIVWFLIDNVFNIVNAAIWPTDN

IDSWWSGVGLCDVEVKLMVASYVGVPGALMCIFRTLAHLLDTDSVTLVPSRGQCRRRLAFEILFCL

VVPAISMVTHYVVQKNRFMLYTIAGCINSFDESWVTLVLSFIWPTVICLTAAYYCCMVLFRICKYK

NQFGTILKSSDSSLSKSRFLRLFFLAFLTLLVILPAQACVLYTNVALSLPWHRYSWGNIHGPSWYT

ITKVATHGQAFFDRWIPIAGGILLFFFFGFGTDATMMYRSFLLRLGLGRCFPDLSRPRNCSARRMS

SISVGPVESFSSRAKTLFSRTQAMVSRTNSIATDPFTPSSSSDIEKGFPISNGSINTIPHRHQLWG

LFTPNGQRNRNMSVPHLPGSENTVSTTAWAGKDQRRSSETDAETIPSSVKNDIIKVKQMIRQTSEK

R*
```

T. crustaceus, preB, α-factor pheromone receptor

Gene sequence:

```
ATGGTCCCTATTGACCCATATACTCAGAATGTTACTCTCCACCGCGCTGATGGCACTCCTTTAGAG

GTGCCTGTGGGCCAGGTGGACTACTTCGTCCAGTATGGCATCAAGATCTGCATCAATTACGGCTCG

CAGATCGGTGCTTCCATCATCCTCCTCATCGTCCTGCTAATGTTGACCAAGCCTGAGAAGCGCTAT

TCCCCCGTGTTCTTCCTGAACAGCAGCGCTCTCGTCTTCAACATCTGCCGGCTGATCTGCATGGCC

ACGTTTTTCACCACCGAGTTTTGCAAGGCCTATCCTTTCCTCTCGAACGACTTTTCAAGCGTTCCA

CTCAGTGCCTACATCAACTCCGTCCTGGGCGTTATCCTGACATTCCTCCTGCTGATCTGCATTGAG

ATCTCCTTTATCCTACAGACCCAGGTCATCTGTTCCACCCTCCGCAACGTCTACCAACGCATCATC

TTCTGGGCATCGATGCTTATGGCTCTCGTTCCAATCGGATTCCGTTTCGCGTTTGTCATCCAAAAC

TCAATACTCATCATGGGAGCAGAGGCCTTTACTTCATATACCTGGCTGCAGAGCGCCACAAATATC

GTCATAACCATCAGCATCTGCTTCTTCAGCGCGATCTTCGTCGTCAAACTGGGCTACACTATTCGG
```

```
CAGCGCAGGAAGCTTGGGGTGAGGAAGTTTGGTCCAATGCAGGTCATCTTCATTATGGGCTGTCAG

ACCATGATCGTTCCAGGTAAGCATAGTATTCCGTGCGTATGCATAGGCAAACTTGGCTGACCTTGC

CAGCTTCTTTCTCAATCATGCAGTATTATATCCCCGTGCCTGAAATGGAATCAAACGTTCTAACGC

TCGTCGCGATATCTCTTCCGCTTTCCTCAATGTGGGCTGCGTCTGCTCTTCACAACCAATCTGAAT

CAGCAGGCGATTCCATCGAACGCCGTAAGCTTTGGAACCCATTTTCATCTTTCAGCCTCGATACGA

GCAGACGGTGGTTGATATTCAAGAATTCCTCTTCGACCGTCACCCACATCAGCAGCGTCTCTTGCT

CGAACTCGACTATGCTTGACCGTCTCTACCCTGACGTGGAGGCTGGAAAGGGGATAGGAGTACACC

GTGACTTCTCAGTGTCCAACTCGCACTGCAAAGAAGTTCTGAATGCGTGA
```

CDS annotation:

```
join(1..742,796..1172)
```

Protein:

```
MVPIDPYTQNVTLHRADGTPLEVPVGQVDYFVQYGIKICINYGSQIGASIILLIVLLMLTKPEKRY

SPVFFLNSSALVFNICRLICMATFFTTEFCKAYPFLSNDFSSVPLSAYINSVLGVILTFLLLICIE

ISFILQTQVICSTLRNVYQRIIFWASMLMALVPIGFRFAFVIQNSILIMGAEAFTSYTWLQSATNI

VITISICFFSAIFVVKLGYTIRQRRKLGVRKFGPMQVIFIMGCQTMIVPASFSIMQYYIPVPEMES

NVLTLVAISLPLSSMWAASALHNQSESAGDSIERRKLWNPFSSFSLDTSRRWLIFKNSSSTVTHIS

SVSCSNSTMLDRLYPDVEAGKGIGVHRDFSVSNSHCKEVLNA*
```