Supporting online material for

Sex determination in the social amoeba *Dictyostelium discoideum*

Gareth Bloomfield*, Jason Skelton, Alasdair Ivens, Yoshimasa Tanaka, and Robert R. Kay

* To whom correspondence should be addressed. Email: garethb@mrc-lmb.cam.ac.uk

This PDF file includes:

Materials and Methods

Growth of *Dictyostelium* strains and macrocyst production

Amoebae were maintained either on SM agar plates in association with *Klebsiella aerogenes* or in axenic medium at 22°C (*S1*). All strains used in this study are described in table S1. To produce macrocysts, cells of different strains to be crossed were grown separately in suspension with heat-killed *K. aerogenes* (5×10^9 bacteria per ml in MSS buffer: 5 mM MES pH 6.0, 10 mM NaCl, 10 mM KCl, 10 mM $CaCl_2$) overnight to a density of $1 – 5 \times 10^6$ cells per ml, washed thrice in fresh MSS and then mixed. For routine crosses 2×10^6 cells of each strain were mixed with 1×10^9 heat-killed bacteria in 2 ml of MSS per well in 6-well plates (Costar, tissue culture treated), before being incubated in the dark for seven days. To quantify macrocyst production in different crosses, 5×10^5 cells of each strain were mixed together with 2.5×10^8 heat-killed bacteria in 0.5ml of MSS in 24-well plates. The plates were then incubated in the dark at 22°C for eight days. In these conditions variable numbers of single-walled cyst-like aggregates are formed as well as true macrocysts; only cysts with clear outer and inner walls are counted, and if multiple cysts form within a single outer wall, all of these are counted (as each is presumably a separate zygote). Strains containing overexpression constructs were grown in the presence of the relevant antibiotic selection during the overnight suspension growth step, but antibiotic was omitted after plating into tissue-culture plates because the wild-type tester strains lack the selectable markers; for consistency in crosses between strains bearing the same markers the antibiotic was still omitted at this step. Macrocysts or cell aggregates were imaged using a Zeiss 710 LSM confocal microscope. Macrocysts can also be readily made from the homothallic strain AC4 by washing amoebae free of food bacteria, resuspending at a density of 1.5×10^7 cells/ml in MSS in Erlenmeyer flasks and shaking at 180 rpm at 22°C in the dark.

Nucleic acid purification and reverse transcription

Genomic DNA was prepared according to (*S2*), or for small scale preparations using GenElute genomic DNA purification columns (Sigma). V12M2 total RNA was extracted by phenol extraction (*S3*). To make *matD* cDNA, a reverse transcription reaction was carried out using random hexamers as primers (Superscript III, Invitrogen). The reaction was carried out at 50°C for 2 hours, before RNase H treatment at 37°C for 20 minutes.

Microarray data

The array experiment analysed (*S2*) is available in ArrayExpress under the accession number E-TABM-394 and the layout of the microarray has the accession A-SGRP-03. The analysis was carried out on data preprocessed using limma (*S4*) in R (*S5*). Heatmaps were prepared using the gplots package (*S6*) in R.

Disruption of the *matA* gene

A region upstream of *matA* was amplified using the primers 5'-AGTGGTACCATAACACCACCAATAGCAGGATTTGAAG-3' and 5'-GGAGAACTCAATTTATAGGTTGTATTGTGC-3' and cloned into the *Kpn*I and *Hind*III sites of pLPBLP (*S7*); a region downstream, including the final 28 nucleotides of the ORF, was amplified using the primers 5'-ATCTATGGAGGCTAAATCTATGACTTGG-3' and 5'-AACCATTGCAATGTCTTTGCTCTACAACC-3' and cloned into the *Bam*HI and *Not*I sites of the same vector. The plasmid was digested with *Kpn*I and *Not*I and purified of protein by phenol-chloroform extraction, then Ax2 cells were transformed by electroporation (*S8*) and selection using 10 µg/ml blasticidin S was maintained until the growth of transformants became visible. Transformants were cloned and screened for the disruption by PCR testing for the insertion of the blasticidin S deaminase at the *mat* locus and loss of the *matA* sequence.

Amplification and sequencing of other versions of the mating locus

The primers D2F (5'-GGTGGTGGTGAACTTAGTAG-3') and C9635 (5'-CAGGATCAGAAATAATTTCTTC-3') in the coding regions up- and downstream flanking genes was used to amplify the locus from good-quality genomic DNA. Successful amplification required the use of the Expand 20kb plus polymerase blend (Roche) and 4.75 mM $MgCl_2$. The cycling conditions were: 92°C 2 minutes, followed by 9 cycles of 92°C for 30 seconds, 52°C for 30 seconds, 68°C for 12 minutes, then 25 cycles of 92°C for 30 seconds, 51°C for 30 seconds, 68°C for 12 minutes plus an additional 20 seconds incrementally each cycle. The NC66.2 locus PCR product pool was partially digested using *Tsp*509I for 30 minutes at 65°C and fragments in the 250-1250 bp range cloned in pBluescript II that had been linearised using *Eco*RI and dephosphorylated using Antarctic Phosphatase (all enzymes from NEB). Inserts were sequenced from both ends, and then assembled. Subsequently versions of the locus were cloned into the vector pSC-A using the

Strataclone PCR cloning kit (Stratagene) and transposon insert libraries (Template Generation System II, Finnzymes) sequenced and assembled. Sequences were later refined in some cases using targeted resequencing of subsequences amplified using the Phusion high-fidelity polymerase (Finnzymes).

Constructs for overexpression of mat genes

MatA- and MatB-GFP constructs were prepared by cloning the respective coding sequences into the *Eco*RI site of pDEX-RH (*S9*) into which *GFP*(S65T) had previously been inserted at the *Hind*III site. The *matC* coding sequence plus stop codon was inserted into the *Eco*RI and *Eco*RV sites of plasmid 339-3 (which contains *mRFP-mars* under the control of the *actin15* promoter in the pBsrH vector backbone (*S10*)). The untagged *matD* expression construct was assembled in three pieces: part of the *matD* cDNA was amplified from cDNA from strain V12M2 using the primers NruF (5'-ATTATTTCGCGAGATGCAACCAATTAC-3') and XhoR (5'-TTTTCTCGAGTTAATACTATAGAACTATTTGAA-3'), cloned into pSC-B (Strataclone kit, Stratagene) then the 5' end of the gene amplified from genomic DNA using the primers using the primers Sac5 (5'-GAGCTCATGAAATTTTTTATTTTTGTTTTTTTCATTT-3') and NruR (5'-CATCTCGCGAAATAATTAAAGGTGATGAA-3') and inserted into the *Sac*I site of pSC-B and the introduced *Nru*I site contained in the NruX primers (this silently changes an *Xba*I site in *matD* to *Nru*I) to make an intermediate construct. The 3' end of the gene was next amplified from genomic DNA with the primers XhoF (5'-TTAACTCGAGAAAAGGAATATTTGCATGATGA-3') and Xba3 (5'-TCTAGATTAAATTAATAAATAAATAATTAAAGGAAATAATAATAAAG-3') and this fragment inserted into the intermediate construct after both were digested with *Xho*I and *Xba*I. This introduces an *Xho*I site into the *matD* coding sequence without altering the polypeptide sequence. The resulting plasmid was then digested with *Sac*I and *Xba*I and the modified *matD* sequence inserted into the pDXA-3H-hygro backbone (*S11*). To make the *RFP-matS* construct the coding sequence plus stop codon of *matS* was amplified and inserted between the *Bgl*II and *Spe*I sites of pDM318 (*S12*). After electroporation of *D. discoideum* cells (see above), transformants were selected with 20 µg/ml G418 (*matA, matB,* and *matS* constructs), 10 µg/ml blasticidin S (*matC* construct), or 40 µg/ml hygromycin (*matD* construct).

Generation of type II locus constructs

A construct containing the entire type II version of the *mat* locus from strain NC66.2 was prepared

by amplifying segments using Phusion DNA polymerase and ligating sequentially into pLPBLP. On one side of the *Bsr* cassette one flanking region amplified with the primers HindF (5'-TTTTAAGCTTTTTCCCAATTTTTTGGTTGTTTTGG-3') and KpnR (5'-GTACTTGGTACCGCAAATGTTGATATAACATC-3') was inserted into the *Hind*III and *Kpn*I sites of pLPBLP. On the other side of the cassette between the *Not*I and *Pst*I sites, three fragments were inserted sequentially after amplification with the primer pairs NotF (5'-AAGTGCGGCCGCACCTGCACCTTCAAAATATA-3') and XbaR (5'-ATTTATCTCTAGATAAAAAGTAAGGAGAG-3'), XbaF (5'-GTATATCTAGATAATAGTGGTGATGTCTT-3') and BamR (5'-TGAAGGATCCATCAATTGAGATATATAAAG-3'), BamF (5'-GATGGATCCTTCAAACCAATTTCAGATTTTA-3') and PstR (5'-TTTTCTGCAGAACAACCAAAAAATTGGGAAAAATCT-3'). This construct incorporates an *Xba*I site into one flanking gene (DDB_G0289171) and a *Bam*HI site into *matD*, neither of which alters the amino-acid sequence. This construct when transformed into the *matA* null strain resulted in strain HM1559, which displays mating behaviour identical to wild-type type II strains according to our assays. An earlier strain, HM1528, which makes abnormal cyst-like structures when plated in the absence of any partner strain after being starved immediately after growth on a bacterial lawn on SM agar plates, originated from a transformation using a different construct. This was prepared by cloning the NC66.2 version of the locus into pSC-A followed by digestion with *Hind*III and *Bgl*II (which cuts 435 bp downstream of *matD*), and the blasticidin S resistance cassette from pLPBLP was then cut with *Bam*HI and *Hind*III and ligated in. A region further downstream of the *Bgl*II site, amplified with the primers 5'-ACAAAAGCTTAATAGATCTCTATACTTTTATT-3' and 5'-GTACTTGGTACCGCAAATGTTGATATAACATC-3' was then inserted into the construct's *Hind*III and *Kpn*I sites. These constructs were transformed into the *matA* null strain after being linearised using *Not*I and *Kpn*I, followed by selection with blasticidin S as above.

Constructs deleting sequentially the *matD* and *matC* coding sequences were based on this modular construct. To delete *matD*, the XbaF-BamR and BamF-PstR were replaced with a fragment amplified with XbaF and PstdelD (5'-AAACTGCAGCATCTTTTTTTTTTAGAAATCTAAAAACTGT-3'); to delete *matB* and *matD* the same two segments were replaced by a fragment amplified with XbaF and PstdelBD (5'-CCACTCGAGTTTCATTAAAATGTTTATTCGGATG-3').

Supplemental discussion

S1. Nomenclature

The mating types of *Dictyostelium discoideum* have been given different names by different authors, and the third mating type has most often treated as a separate class ('bisexual' or 'ambisexual'; on the basis of available evidence we see no reason to treat these isolates in any way other than a third gender since each of the three known genders can mate with both of the other two, but not itself). Clark *et al* (*S13*) used 'type I' and 'type II' for the mating types exemplified by the NC4 and V12 isolates respectively; then Erdos *et al* (*S14*) used *matA1* and *matA2* (later Urushihara and Muramoto (*S15*) suggested extending this to include *matA3*); finally Robson and Williams (*S16*) suggested *matA* and *mata*. We consider that because the mating types do not vary as simple alleles – no gene is common to all versions of the locus – the scheme of Erdos *et al* is not adequate, and because the *A/a* system is not easily extendible beyond two mating types, neither is that of Robson and Williams. We therefore have reverted to the scheme of Clark *et al*., adding type III. We note the irony that type II, having a version of the locus that effectively combines the other two, better fits the designation of 'bisexual:' it had previously been conjectured that type III cells were contained both type I and II versions of the locus (*S14*, *S17*).

In naming genes and proteins, we follow the scheme of Demerec *et al* (*S18*, *S19*). We believe the locus we describe is the same as that previously studied (*S16*, *S17*), and are confident in following them in giving it the name *mat.* We describe the different versions of the locus in different mating types as idiomorphs, not alleles, following the terminology used to describe fungal sex-determining sequences which differ in gene content (*S20*). We treat the genes that are homologous but in different idiomorphs to be different genes (thus *matA* and *matB* are homologous genes derived from a single ancestral sequence, as are the *matC/matS* and *matD/matT* pairs, while because the type III and homothallic versions are clearly more closely related and do not differ obviously in gene content, we consider the homologous sequences within them to be allelic and thus name these genes *matS* and *matT* in both. Similarly because the locus in *Dictyostelium purpureum* and *Acytostelium subglobosum* appears to have the same structure as the type III version, we call the genes in the sequenced isolates of these species *matS* and *matT*. Finally although we note that *matD* and *matT* do not appear to be involved in mating-type determination, we retain the *mat* designation for them because they clearly form part of the locus defined first by the *matA* gene, and additionally seem likely to function during the process of mating (broadly

construed).

S2. Previous attempts to map the mat locus by linkage analysis. The locus we have identified lies between DDB_G0289163 and DDB_G0289171, which were mapped to chromosome 5 (previously referred to as linkage group VII in early pre-genome genetic studies) during the sequencing of the genome (*S21*). The first attempt to map the mating type locus (*S22*) used sexual crosses of strains resistant to cycloheximide and methanol. No evidence of linkage of the mating locus to the genes conferring resistance to these compounds was found, and since only single genes are known that confer resistance to each are known, on chromosomes 1 and 2 respectively, this suggested that *mat* must lie on one of other chromosomes. Later work on progeny from 'escaped' diploids from parasexual crosses between type I and II haploids (*S23*) gave conflicting evidence: these authors noted that in several crosses, diploids that had become homozygous at the mating locus also had become homozygous at a spore shape locus (*sprA*) on chromosome 1 (C1), but not at the linked *cycA* and *tsgE* loci. The most parsimonious explanation in the absence of other evidence was that *mat* was also to be found closely linked to *sprA* on C1, distal to *tsgE*. However the diploids in question were not usefully marked on chromosome 5 (C5). It is difficult to reconcile these two earlier studies, or the second study with our discovery of a locus bearing all the expected properties of the mating locus on C5. Although it is possible that all of Welker and Williams' strains had undergone multiple recombination events leading to loss of heterozygosity at loci on both C1 and C5, perhaps the best explanation lies in complications regarding spore shape: the *sprA* mutation has been reported to be dominant (*S24*), and wild-type V12 bears at least one locus affecting spore morphology (*S25*) so perhaps complex interactions between the multiple loci from two genetic backgrounds that affect this phenotype made scoring based on gross morphology problematic.

S3 Mating type switch from type I to type II. The conversion of our laboratory type I strain Ax2 to type II by the transfer of the entire genomic sequence of the type II version of the locus into the *matA* null strain was ultimately successful. Strain HM1559, which contains *matB*, *matC* and *matD* in their native order and with their regulatory sequences intact, but not integrated *in situ* at the *mat* locus, mates qualitatively in the same way as wildtype type II strains (as far as our assays can determine). One earlier attempt, using a different construct, produced a more complex phenotype. This strain, HM1528, behaved as a normal type II after standard conditions of growth (fed on heat-killed bacteria in shaken suspension in the dark), but when taken from less-controlled growth

conditions (from the growing edge of a plaque forming on a lawn of live bacteria) this strain produced abnormal cyst-like structures when plated in submerged non-nutrient buffer alone in the absence of a type I or type III partner. This abnormality is not understood, but presumably reflects a mutation either in the introduced *mat* sequence or elsewhere in the genome; again this construct had not replaced the deleted *mat* locus, so it is possible that the insertion site of the construct is of importance.

S4. Homothallic isolates. The genetic difference underlying the distinct type III and homothallic phenotypes is still under investigation. It is possible that the differences in sequence between the type III and homothallic *mat* coding sequences are alone responsible for self-fertility (and its absence in type III), or one or more further loci may be required. This is a continuing focus of our work, as we believe that the functional and evolutionary links between these strains will prove to be illuminating. Whichever of these alternatives is true, this question does not affect the status of *mat* as the single locus determining sex in *D. discoideum*: the ability to undergo autogamy (or apogamy) is a separate phenomenon although clearly it also falls under the more general category of mating behaviour.

S5. Relationship with other sex-determining systems. The organisation of the *Dictyostelium* mating-type locus does not perfectly resemble previously studied sex-determining regions, although its compactness and the complete dissimilarity between different versions is reminiscent of those in certain fungi *(S20)*. To our knowledge the threefold organisation of homologous sequences is unique to this system. It does not appear to be closely similar to the known multiallelic mating-type loci of the ciliate *Euplotes raikovi (S26)* or of basidiomycete fungi *(S27)*. These rely at least in part on variable pheromone and receptor genes that prevent self-mating. Basidiomycetes typically possess a two-locus system, in which one of the two loci responsible encodes pairs of homeodomain transcription factors that discriminate between self and non-self through the ability to heterodimerize *(S28)*. This general kind of recognition event is also found in incompatibility systems that prevent self-fertilization in many flowering plants (*S29*) and that underlie adaptive immune responses in animals (*S30*). Until the molecular interactions between *Dictyostelium mat* genes are understood we cannot know to what extent convergent similarities between these systems have evolved. However we believe that we can make a distinction between the mating-type system in *Dictyostelium* (and also those in ascomycetes and the isogamous algae) which have a fixed, stable number of compatibility groups (sexes) and these other incompatibility systems which appear

to be evolving much more rapidly under pressure to generate new forms. We have sampled and sequenced multiple independent isolates of each *D. discoideum* mating-type, and find almost perfect conservation of all the *mat* genes; the mating ability of dozens of other isolates have been determined by other groups, finding all three mating types in equivalent numbers in many locations,and no additional mating-types have been found. Thus it seems that the evolutionary history of this species has fixed three stable sexes in this species. *Why* this happened has to remain an open question.

Legends to supplemental tables

Table S1. Strains used in the present study. Strains, their parents, mating type, overall known genotype and source are listed where known. If no parent is listed, the strain is a wild isolate, a wild-type laboratory line, or in the case of A2cycR a mutant whose parental strain is not clear from the literature. Mating-type 'IIb' is defined here as the partial type II phenotype of strains containing *matB* but not *matC*; these strains will form macrocysts when paired with type III cells, but not with type I nor type III. Similarly, type IIc is defined as the partial type II phenotype resulting in strains expressing *matC* but not *matB*; these strains form macrocysts with type I but not type III nor type II. Some strains generated in the present study are listed with a mating-type as '-' because although they do not form macrocysts in our assays it is possible that they retain some mating-type specific function that we have not tested for and so it is not certain that they are strictly *mat* null like their immediate parent. Sources are as follows: A – described in (*S31*), obtained from the Dicty Stock Center (http://www.dictybase.org/StockCenter/StockCenter.html)**;** B – *(S22, S32*), obtained from David Francis; C – this laboratory; D – a line of NC4 *(S33)* obtained from Pauline Schaap; E - *(S34)* obtained from David Francis; F – Japanese isolate obtained from Hiromitsu Hagiwara; G – line of V12 isolate obtained from Gunther Gerisch; H – described in *(S14),* and obtained from the Dicty Stock Center; I – described in *(S17)* and obtained from the Dicty Stock Center.

Table S2. Genes missing or divergent in one or more wild isolate. Data are given for all genes which gave a logratio below -1 in one or more strains when compared in array comparative genomic hybridisations with the reference Ax2 strain. The mating type of each strain is given in parentheses in the column headers. More genes are presented here than in figures 2 and S2, for which a cutoff of -2 was used to make visualisation clearer. A single gene, DDB_G0289165 (*matA*), has logratios below -1 for all type II strains and logratios above -1 in all type I strains.

Table S3. Conservation of *matA* coding sequences in type I isolates. The percent identities of the nucleotide and amino-acid sequences for three type I wild isolates are given, all compared to the Ax4 sequence that is found in the genome databases. The sequence in Ax2 is identical (Ax2 and Ax4 are related laboratory strains deriving from the same wild isolate, NC4). In WS205 there is one silent nucleotide change at base 225. The sequence of the *mat* locus of WS205 is available in the EMBL database *(S35)* under the accession FN543123. The Ax4 *matA* gene is annotated as such in dictyBase *(S36).*

Table S4. Conservation of *matB, matC,* and *matD* coding sequences in type II isolates. The percent identity of the nucleotide sequences of each *mat* gene is shown compared to those from the first version of the locus to be sequenced, NC66.2. Where the identity is less than 100%, the absolute number of nucleotide and amino-acid changes are given in parentheses. The sequences of the *mat* locus from strains NC66.2 and NYA64 are available in the EMBL database under the accessions FN543124 and FN543121.

Table S5. Re-engineering of the mating behaviour of a type I strain, full results. Crosses were performed as described in the legend to figure 3; briefly, pairwise crosses were made between wild-type strains, the *matA* null strain, strains expressing in various combinations the genes located in the three versions of the *mat* locus (with the exception of *matT*), two strains in which the *matC* or both *matB* and *matC* genomic sequences replaced the knocked-out version of the locus (marked 'C gen' and 'BC gen' respectively), and finally a strain containing the entire genomic sequence of the type II locus integrated elsewhere in the *matA* null genome (marked 'BCD gen') The wild-type strains are Ax2 (WT-I), V12M2 (WT-II), and WS2162 (WT-III). In three independent crosses in each case, macrocysts were counted after eight days. The mean and standard error (in brackets) are shown. ND = not determined; we have no reason to believe that the *matA* expressing strain differs in its mating behaviour from wild-type type I strains, which only possess *matA* at the mating locus.

Table S6. Conservation of *matS* and *matT* coding sequences in type III and homothallic isolates. Percent identities of the *matS* and *matT* nucleotide and amino-acid sequences are given comparing the type III isolate WS112B and the homothallic isolates AC4 and ZA3A with the type III WS2162. The nucleotide sequences of the *mat* locus from strains AC4 and WS2162, and the *matS* and *matT* genes of ZA3A are available in the EMBL database under accessions FN543120, FN543122, FR666792, and FR666793 respectively.

Table S1.

Strain	Parent	Mating type	Genotype	Source
AC4		homothallic	*matS2, matT2*	A
A2cycR		II	*matB4, matC4, matD4*	B
Ax2	Ax1 (NC4 derived)	I	*matA1*	C
HM1524	Ax2	null	*matA-*, bsR	C
HM1526	HM1524	null	*matA-*	C
HM1528	HM1526	II	*matA-, matB1, matC1, matD1,* bsR	C
HM1555	HM1526	II	*matA-, matB1, matC1,* bsR	C
HM1559	HM1526	II	*matA-, matB1, matC1, matD1,* bsR	C
HM1560	HM1526	IIc	*matA-, matC1,* bsR	C
HM2883	HM1526	IIb	*matA-, [act15]:matB1:GFP,* neoR	C
HM2885	HM2883	II	*matA-, [act15]:matB1:GFP, [act15]:RFP:matC1,* neoR, bsR	C
HM2886	HM1526	IIc	*matA-, [act15]:RFP:matC1*, bsR	C
HM2899	HM2886	IIc	*matA-, [act15]:RFP:matC1*, *[act15]:matD1,* bsR, hygR	C
HM2902	HM2885	II	*matA-, [act15]:matB1:GFP, [act15]:RFP:matC1*, *[act15]:matD1,* bsR, hygR, neoR	C
HM2906	HM1526	-	*matA-,[act15]:matD1,* hygR	C
HM2920	HM2883	IIb	*matA-, [act15]:matB1:GFP, [act15]:matD1,* hygR, neoR	C
HM2930	HM1526	III	*matA-, [act15]:RFP:matS1,* neoR	C
HM2960	HM1526	I	*matA-/[act15]:matA:GFP,* neoR	C
NC4(S)		I	*matA1*	D
NC28.2		I	*matA3*	E
NC42.1		II	*matB4, matC4, matD4*	E
NC59.2		I	*matA4*	E
NC66.2		II	*matB1, matC1, matD1*	E

NC94.2		II	*matB5, matC5, matD5*	E
NYA64		II	*matB3, matC3, matD3*	F
V12M2		II	*matB2, matC2, matD2*	G
WS112B		III	*matS3, matT3*	H
WS205		I	*matA2*	B
WS2162		III	*matS1, matT1*	H
ZA3A		homothallic	*matS4, matT4*	I

Table S2

Gene ID	NC4S (I)	NC28.2 (I)	NC59.2 (I)	WS205 (I)	A2cycR (II)	NC42.1 (II)	NC66.2 (II)	NC94.2 (II)	NYA64 (II)	V12M2 (II)
DDB_G0268490	0.16	0.24	0.18	-0.82	-0.23	-0.11	0	-1.01	0.86	-0.9
DDB_G0268498	-0.29	-0.36	-0.16	0.25	0.07	-1.27	0.04	0.05	0.07	0.64
DDB_G0268522	0.07	0.01	0	-0.12	0.05	-0.06	-1.09	0.11	-0.43	0.5
DDB_G0267902	0.99	0.26	0.37	-0.11	1.19	0.24	0.33	0.91	-1.28	0.86
DDB_G0268012	0.38	-0.05	0.11	-1.13	-0.82	0.03	0.03	0.07	-0.63	0.18
DDB_G0268132	-0.05	0.87	-1.06	0.08	0.05	0.24	-0.08	0.06	0.06	0.16
DDB_G0269010	0.05	0.04	0.23	-0.1	-0.04	-0.11	0.31	-0.78	-1.14	0.45
DDB_G0269408	-0.02	0.36	0.04	0.54	0.48	0.04	0.09	0.1	-2.2	0.55
DDB_G0270514	0.15	0.33	-0.25	0.06	0.06	-1.08	-0.07	-0.08	-0.95	0
DDB_G0270598	0.11	-0.68	-0.52	-0.15	-0.16	-0.61	-0.54	-2.21	-0.35	0.08
DDB_G0270600	0.05	-0.08	-0.16	0.09	-0.09	0	0.28	0.03	-1.21	0.02
DDB_G0269226	0	-0.17	-0.15	0.12	0.03	-1.04	-0.14	-0.11	-0.01	0.19
DDB_G0270970	0.17	0.02	0.13	-0.61	-0.03	-0.01	-0.17	-0.01	-2.46	0.01
DDB_G0269818	-0.13	-0.44	-0.09	-1.02	-0.5	-0.28	-0.14	-0.13	-0.12	-0.03
DDB_G0269870	-0.16	-0.07	-0.13	-1.16	-0.11	-0.2	-0.25	0.29	0.05	0.18
DDB_G0269954	-0.35	-0.02	0.46	0.05	0.1	0.12	-1.02	0.05	0.01	0.27
DDB_G0270696	-0.12	0.17	0.13	-1.47	-0.13	0.06	0.07	0.07	-0.76	-0.11
DDB_G0270744	0.71	0.15	0.24	-0.06	1.34	0.27	0.1	0.99	-1.8	0.51
DDB_G0270256	-0.05	-0.21	-0.04	-1.04	0.02	0.09	0.26	-0.02	-0.06	0.49
DDB_G0270348	-0.05	0.03	0.04	-1.11	-0.57	-0.13	-0.21	0	-0.08	-0.16
DDB_G0270370	0.09	-0.04	0.38	-0.13	0.15	-0.07	-0.05	-0.2	-0.07	-1.18
DDB_G0270400	0.37	0.05	0.1	-2.84	-2	-0.2	-0.07	0.05	-2.18	-0.22
DDB_G0294302	0.01	-1.05	-0.1	-1.01	-1.73	0.14	-0.09	-2.01	-0.6	0.62
DDB_G0294222	0.05	-0.11	0.19	0.2	0.04	0.16	0.07	0.01	-1.26	0.15
DDB_G0294314	0.15	-0.08	0.01	-0.64	0.16	0.19	0.14	0.18	-1.21	0.07
DDB_G0294158	0.11	0.47	-0.16	-3.13	-1.87	0.03	-0.1	-0.09	-2.12	-0.22
DDB_G0271160	-0.05	0.12	-0.07	-0.27	-2.45	0.28	-0.15	-0.4	0.06	-0.26
DDB_G0271158	0.17	0.08	-1.96	0.23	0.21	-0.33	-0.01	0.02	-0.86	0.11
DDB_G0271646	0.05	0.15	-0.17	-1.66	-1.19	-0.1	-0.05	-0.27	-0.29	-1.92
DDB_G0271650	0.12	0.03	-0.01	-1.84	-1.57	-0.05	-0.05	-0.11	0.13	-2.34
DDB_G0272428	-0.17	0.3	-0.08	-0.79	-0.37	0.08	-0.35	-0.05	-1.02	-0.01
DDB_G0272164	-0.05	-0.02	-0.04	-2.53	-1.64	0.07	-0.02	0.03	-2.82	0.35
DDB_G0272945	0.04	0.04	-0.02	-2.73	-0.38	0.01	-0.02	-1.76	-2.08	-1.85
DDB_G0272714	0.08	-0.11	0.13	-0.11	-1.91	-0.02	0.05	-0.04	-0.15	0.07
DDB_G0272706	0.01	-2.44	-0.05	-3.1	-1.86	-0.08	-0.11	-0.1	-2.75	-2.57
DDB_G0272995	0.06	-1.29	-0.12	0.17	0.13	-0.12	0.05	-0.02	0.12	-0.1
DDB_G0272833	0.01	-0.08	-0.15	-0.41	0.01	-0.02	-0.04	-0.12	-1.1	-0.24
DDB_G0273799	-0.16	0.01	0.06	-0.38	-0.4	0.09	0.04	0.04	-1.7	-0.05
DDB_G0273959	0.05	-0.36	-0.03	-0.42	-0.5	-0.06	0.04	-0.02	-1.03	0.04
DDB_G0273963	0.04	0.09	0.05	0.08	0.2	0	-0.01	-0.01	-1.1	-0.17
DDB_G0274087	0.33	-1.73	0.12	-0.19	-2.64	0.61	-0.36	0.69	-1.27	-0.33

DDB_G0274377	-0.03	0.13	-0.15	-0.12	-0.06	-0.02	0	-0.06	-2.21	0.1
DDB_G0274465	0.1	-0.19	-0.04	-0.47	0.1	-0.1	-0.12	0	-0.36	-2.61
DDB_G0274613	-0.1	-0.07	0.08	-1.54	-0.1	-0.06	0.46	0.78	0.97	0.2
DDB_G0275103	0.09	0.24	0.83	-0.02	0	-1.24	0.82	-0.07	0.03	0.56
DDB_G0275013	0.08	0.01	0.02	-0.28	0.01	-0.21	0.04	-0.07	-2.41	-0.23
DDB_G0275353	0.12	-0.06	0.15	0.22	0.09	0.17	-0.08	0	-2.2	-0.16
DDB_G0275783	0.15	0.23	0.11	-2.21	-1.61	0.05	0.05	0.06	-2.16	-0.25
DDB_G0275659	0.02	0.19	-0.03	0.12	0.16	0	0.08	0	-1.52	0.24
DDB_G0275487	0.09	-0.02	-0.01	-0.02	-0.04	0.08	0.05	-0.04	-1.61	-0.5
DDB_G0276373	-0.06	-0.06	-0.08	-0.29	-0.05	0.15	-0.02	-0.04	-1.51	0.13
DDB_G0276367	0.08	-0.09	0	-0.17	0.16	-0.01	0	0.09	-1.45	0.06
DDB_G0277063	-0.12	0.02	0.01	-0.18	0.28	-0.02	-0.01	-0.08	-1.77	0.05
DDB_G0277083	-0.02	0.23	0.07	-0.29	-0.05	0.01	-0.04	0.12	-2.11	-0.33
DDB_G0277317	-0.09	0.05	-0.07	-1.31	-0.16	0.04	-0.09	0.2	0.35	0.32
DDB_G0277287	0.55	0.21	0.26	-0.19	0.67	0.33	0.32	0.65	-1.13	0.7
DDB_G0277461	0.27	0.02	0.08	-0.13	-0.15	-0.02	0.11	-0.02	-1.11	0.07
DDB_G0277463	-0.09	-0.38	-0.03	-0.04	-0.12	-0.09	0.42	0.02	-1.85	-0.02
DDB_G0277465	0.07	0.06	-0.2	-0.05	0.06	-0.29	-0.19	-0.04	-1.5	0.13
DDB_G0277467	-0.02	-0.08	-0.04	0.15	0.14	-0.14	-0.11	-0.01	-2.2	-0.13
DDB_G0277731	0.03	0.19	-0.01	-0.1	-0.03	0.01	-1.04	-0.05	-0.02	-0.14
DDB_G0277807	0.05	0.09	-0.09	-2.12	0.05	-0.03	0.12	0.06	-0.07	-0.02
DDB_G0295721	-0.07	0.04	0.12	-0.01	-0.22	-0.16	-1.07	0.21	-0.38	0.09
DDB_G0278599	-0.09	0.04	0.12	-1.75	-0.14	0.01	-0.04	-0.06	-0.18	0.01
DDB_G0278133	0.06	0.05	-0.17	-0.12	0	-0.1	-0.06	0.55	-2.49	-0.21
DDB_G0278139	0.25	-0.1	-0.05	-0.33	0.03	-0.06	0.02	0.15	-2	0.02
DDB_G0278177	0.11	-0.28	0.09	0.11	0.11	-0.05	0.14	-0.17	-2.56	-0.48
DDB_G0278303	0.02	-0.15	-0.07	-0.46	0.03	-0.12	-0.02	-0.1	-2.56	-0.29
DDB_G0278667	0.1	-0.01	-0.12	-0.8	-1.24	0.4	0.13	0.31	-0.77	-0.52
DDB_G0278363	-0.04	-0.02	-0.01	-4.1	-3.17	-0.07	-0.11	-0.11	-3.44	-3.4
DDB_G0278525	-0.15	0.37	0.03	-1.21	0.17	0.44	-0.08	-0.03	-0.09	0.19
DDB_G0279071	0.07	0.06	0.01	-0.05	0.21	0.09	0.15	-0.11	-2.92	-0.07
DDB_G0279279	0.06	0.09	0.01	-1.23	-0.04	0.15	-0.19	-0.33	0.15	-0.22
DDB_G0280005	0.11	-2.15	-0.09	0.16	0.01	-0.11	-0.08	0.06	0.17	0.14
DDB_G0280303	0.1	-0.14	0.19	0.05	0.1	-0.1	-0.07	-0.08	-2.26	-0.09
DDB_G0280655	-0.1	-0.93	-1.17	-3.56	-2.63	-1.02	-0.96	-2.21	-2.97	0
DDB_G0280659	-0.1	-4.1	-4.51	-3.94	-0.28	-4.5	-3.69	-3.27	-2.43	-0.04
DDB_G0280661	-0.19	-2.86	-2.9	-4.16	-0.51	-3.14	-2.59	-3	0.38	-0.74
DDB_G0280663	-0.01	-3.04	-2.84	-4.11	-2.38	-3.09	-2.61	-3.85	-2.5	-2.1
DDB_G0280665	0.07	-3.46	-3.19	-4.52	-2.85	-2.97	-2.41	-3	-2.57	-2.35
DDB_G0280667	-0.05	-4.04	-4.28	-4.08	-0.1	-4.03	-3.56	-3.1	-3.12	-0.13
DDB_G0280673	-0.12	-2.52	-2.81	-3.28	-2.93	-2.57	-1.96	-1.95	-2.19	-0.39
DDB_G0280885	-0.09	-0.21	-0.02	-0.02	-0.17	-0.03	0.12	-0.15	-1.1	-0.18
DDB_G0280993	-0.22	-0.14	-0.06	0.13	-0.02	-0.25	-0.01	-1.29	0.01	0.04
DDB_G0281159	-0.14	-0.11	-0.14	-1.31	0.11	-0.02	0.32	-0.2	0.67	-0.15
DDB_G0281371	0.09	-1.41	-0.07	0.4	0.04	0.05	-0.05	0.04	0	0.07

DDB_G0281373	-0.04	0.01	-0.04	0.05	0.12	-0.03	-0.07	-0.01	-1.52	0.01
DDB_G0281417	-0.17	0.27	-0.26	-0.3	0.26	-1.07	-0.11	-0.2	-0.28	0
DDB_G0281571	0.23	-0.09	0.08	-0.03	0.18	-0.29	-1.07	0.04	0.15	-0.09
DDB_G0281645	0.01	-0.08	0.24	0.06	0.04	0.02	0.04	-0.07	-1.74	0.29
DDB_G0281695	-0.03	0.02	0.07	-0.42	-1.11	0.04	-0.45	-0.09	-0.02	-1.35
DDB_G0281959	-0.05	0.03	0.01	0.11	-0.11	0.19	0.12	0.09	-1.63	0.23
DDB_G0282163	-0.09	0.26	-0.08	0.07	-2.87	-0.12	-0.13	0.18	-0.04	0.06
DDB_G0282285	0.2	-2.11	-2.22	0.19	0.17	-0.2	-2.45	-2.34	0.02	0.05
DDB_G0282401	-0.07	-0.18	0.05	-1.13	-0.26	0.16	0.04	0.08	-0.49	0.09
DDB_G0282437	-0.1	0.15	-0.03	-1.44	-1.3	0.02	0.05	0.09	-0.36	-0.11
DDB_G0282503	-0.09	-0.01	0.06	0.17	-0.03	0.28	0.09	0.16	-1.59	0.15
DDB_G0282637	0	0.08	-0.22	-0.28	-0.48	0.09	-0.07	0.31	-1.62	-0.02
DDB_G0282699	0.06	-0.03	0.08	-0.04	0.29	0.01	-0.66	-1.28	-0.02	0.03
DDB_G0282857	0	-0.13	-0.06	-1.65	-0.24	-0.12	-0.22	-0.09	-0.02	0
DDB_G0282877	-0.13	-0.19	-0.11	0.04	-0.12	-0.13	0.04	-0.15	-1.1	-0.3
DDB_G0283211	-0.08	-0.04	-0.06	-0.29	-0.13	-0.1	-0.01	0	-2.66	-0.21
DDB_G0283257	0.03	-1.15	-0.02	0.12	0.14	-1.1	-1.3	-1	-1.22	-1.03
DDB_G0283337	0.01	0	0.01	-3.11	-3.12	0.12	-0.01	-0.01	0	-3.65
DDB_G0283341	0.14	-1.56	-0.06	-2.3	-1.65	0.03	0.04	-1.93	-2.04	-1.48
DDB_G0283429	0.03	-0.08	-0.1	0.09	0.13	0.08	0	0.08	-1.81	-0.11
DDB_G0283997	-0.06	0.1	0.03	0.28	-0.07	0.16	0.04	0.01	-1.86	0.12
DDB_G0284057	0.01	0.05	-0.02	0.32	0.4	-0.17	-0.19	-0.05	-1.4	0.31
DDB_G0284191	-0.08	-0.51	0.02	-0.12	-0.34	-0.21	-0.02	-0.09	-1.06	0.07
DDB_G0284269	0.18	-0.08	0.06	-0.07	0.02	0.04	0.01	0.07	-0.18	-1.73
DDB_G0284271	-0.1	0.04	-0.01	-0.51	-0.28	0.08	0.01	-0.15	-0.03	-1.67
DDB_G0284533	0.03	-0.03	-0.05	0.19	-1.1	0	-0.12	0.06	0.05	-0.11
DDB_G0284469	0.09	0.05	0.07	0.43	0.2	0.12	-0.04	-1.1	0.27	0.29
DDB_G0284763	0.04	0.07	-0.1	-0.22	-0.18	0	-0.04	-0.08	-2.2	-0.05
DDB_G0284763	-0.13	-0.12	-0.08	-0.03	-0.16	0.03	-0.03	-0.04	-2.52	0.01
DDB_G0284859	0.08	0.01	-0.26	0.07	-0.19	-0.29	-0.03	-0.05	-1.85	-0.08
DDB_G0284865	0.06	0.02	-0.05	0.27	0.12	-0.14	0.07	0.07	-1.74	-0.03
DDB_G0285069	-0.3	-0.1	-0.18	-1.32	-0.06	-0.04	-0.07	-0.07	0.27	-0.1
DDB_G0285077	0.07	0.26	0.07	-1.35	0.69	-0.04	0.19	0.11	-0.11	0.77
DDB_G0285437	-0.22	0.02	0.09	0.45	0.04	-0.06	0.44	0.07	-2.12	0.31
DDB_G0285585	-0.09	0.16	0.03	0.4	0.1	0	0.13	0.03	-1.33	0.17
DDB_G0286489	0.03	-0.13	-0.02	-0.07	-0.02	-0.16	-0.08	0.01	-1.53	0.11
DDB_G0286517	0.01	-0.06	0.18	-1.29	0.16	-0.06	-0.14	0.15	-0.3	0.23
DDB_G0286501	0.56	0.16	0.23	-1.35	0.14	0.07	0.71	0.47	0.09	0.33
DDB_G0286853	0.12	-0.2	0.18	-2.02	0.03	-0.16	-0.05	0	-1.49	-2.01
DDB_G0286881	-0.21	-0.13	0.06	-1.14	-0.07	0.24	0.17	0.88	0.75	0.36
DDB_G0287235	-0.06	0.09	-0.02	-0.12	0	0.15	0.06	-0.09	-2.48	-0.11
DDB_G0287885	-0.13	0.14	0.06	0.39	0.04	0.1	0.01	0.04	-2.6	0.01
DDB_G0287969	-0.02	-0.09	-0.08	-1.03	-0.26	-0.07	0.01	0.41	-0.02	-0.13
DDB_G0288253	-0.01	-0.01	-0.02	-0.31	-0.11	-0.03	-0.12	0.06	-0.14	-2.22
DDB_G0288283	0	-0.06	-0.02	-0.04	-0.3	0.13	0.14	0.16	-2.88	-0.05

DDB_G0288417	0.1	-0.09	-0.22	-1.74	-1.29	-0.12	-0.01	0.04	0.68	-1.37
DDB_G0288455	0.12	-0.04	-0.05	0.03	-2.32	-0.02	-0.15	-0.17	-0.08	-0.09
DDB_G0288577	0.05	0.01	-0.02	-1.84	-3.07	0.06	0.05	0.11	-1.54	-0.83
DDB_G0288579	-0.07	-0.12	0.04	-4.27	-3.24	0.08	0.06	-0.05	-2.6	-2.11
DDB_G0288581	0	0.13	0.06	-3.2	-2.47	0.21	0.18	0.09	-2.61	-2.81
DDB_G0288587	0.11	0.2	0.14	-2.16	-1.72	0.06	-0.07	0.15	-2.31	-1.16
DDB_G0288683	0	0.07	0.06	0.15	-0.03	-0.05	-1.29	0.04	0.09	-0.01
DDB_G0288685	-0.03	0.05	0.02	0	0.03	-0.05	-1.11	-0.1	-0.14	-0.07
DDB_G0288767	-0.02	0	0.02	-0.06	-0.03	-0.27	0.07	0.11	-1.8	0.08
DDB_G0289081	-0.01	-0.03	-0.16	0.26	0.09	-0.17	-1.05	-0.03	-0.07	0.09
DDB_G0289165	0.1	0.44	0.11	-0.05	-1.7	-2.28	-1.77	-2.31	-2.39	-1.57
DDB_G0289227	0.09	0.11	0.01	-1.09	0.08	0	0.02	-0.01	0.05	-0.25
DDB_G0289173	-0.07	-0.02	-0.01	-1.17	-0.19	-0.15	-0.13	-0.05	0.06	-0.16
DDB_G0289235	0	-0.04	0.01	-1.17	-0.19	-0.1	-0.05	-0.03	0.07	-0.29
DDB_G0289401	0.07	-0.03	-0.05	-1.12	-0.59	-1.08	-0.13	-0.89	-0.83	-0.67
DDB_G0289407	0.08	0.09	0.06	-1.11	-0.79	-0.8	0.04	-0.85	-1.12	-1.05
DDB_G0289425	0.06	0.04	0.14	-0.1	0.06	0.5	-0.07	0.37	-1.09	-0.58
DDB_G0289529	-0.02	0.18	-0.02	-1.24	0.12	-0.13	-0.27	0.01	-0.06	-0.05
DDB_G0289679	-0.1	0.21	0	-1.31	-0.06	0.36	0.52	0.85	0.88	0.55
DDB_G0289731	0.16	0.74	-0.05	0.24	-0.49	0.34	0	0.23	-1.25	0.42
DDB_G0290101	-0.05	0.15	0.04	0.12	-1.69	-2.22	-0.04	-0.3	-2.64	0.06
DDB_G0290251	0.02	0.08	0.08	-1.2	0.12	-0.04	-0.72	0	-0.03	-0.16
DDB_G0290557	-0.06	-0.01	-0.09	-1.13	-0.3	-0.06	-0.05	0.01	-0.02	-0.07
DDB_G0290679	-0.02	0.61	0.1	-0.1	-0.09	0.48	0.41	0.49	-2.19	0.05
DDB_G0290607	-0.02	0.23	0.1	-0.19	-2.95	0.12	-0.13	0.29	-0.43	-0.65
DDB_G0291610	0.02	0.02	-0.11	-1.17	-0.2	-0.07	-0.04	-0.07	-0.1	0.04
DDB_G0291678	0.06	0.06	0.26	-0.05	0.12	0	0.06	0.14	-1.87	-0.05
DDB_G0291774	0.03	-0.25	-0.21	-2.57	-1.68	-0.07	-0.12	-0.92	-1.25	-1.78
DDB_G0291946	0.12	-0.98	0.04	-0.6	-0.05	-0.01	-0.55	0.21	-1.08	-0.81
DDB_G0292182	0.01	-2.03	0.17	-0.2	-1.77	0.2	0.09	0.13	-2.65	-1.97
DDB_G0292280	0.16	-0.25	0.05	-0.19	-0.21	-0.03	-0.07	-1.02	-0.07	-0.5
DDB_G0292354	-0.35	-0.05	-0.01	-0.24	-1.11	-0.13	-0.14	-0.09	0.06	-0.23
DDB_G0292532	0.01	-0.03	0.01	0.07	0.41	0.03	-0.03	-0.03	-1.99	0.18
DDB_G0292606	0.08	-0.05	-0.1	-0.03	0.18	-0.02	-0.14	0.03	-2.59	0.05
DDB_G0292842	-0.11	-0.2	-0.02	-0.7	-0.23	0.06	-0.3	-0.05	-1.97	0.02
DDB_G0292844	-0.07	-0.02	-0.04	-0.49	-0.34	-0.02	0.01	-0.09	-2.19	-0.26
DDB_G0292962	-0.16	-0.06	0.1	-0.39	-0.06	-0.02	-0.01	0.05	-1.8	-0.21
DDB_G0292908	-0.03	-0.07	0.09	0.03	-0.1	-0.17	-0.03	-0.07	-1.48	0
DDB_G0293020	-0.01	0.1	0.61	0.02	0.45	0.23	0.48	-0.02	-2.63	0.27
DDB_G0293336	-0.06	0.03	-0.07	-1.08	-0.19	-0.09	-0.08	-0.06	0.06	-0.1
DDB_G0293336	0	-0.1	-0.18	-1.03	-0.14	0	0.06	-0.03	0.11	-0.12
DDB_G0293424	-0.06	-4.03	-3.89	-3.57	-2.56	-3.99	-3.3	-2.88	-2.97	-4.28
DDB_G0293430	0.02	-0.64	-0.47	-1.8	-1.79	-0.34	-0.26	-1.4	-0.59	-0.7
DDB_G0293464	-0.07	-2.57	-2.71	-2.62	-1.6	-2.63	-1.71	-2.38	-2.5	-2.88
DDB_G0293456	0.11	-2.24	-2.76	-3.28	-2.36	-2.14	-2.4	-1.9	-3.03	-2.01

DDB_G0293456	0.01	-2.06	-2.06	-2.82	-2.12	-2.03	-0.81	-2.5	-3.11	-1.49
DDB_G0293528	-0.08	0.07	-0.24	-1.09	-0.61	0.25	0.25	0.3	-0.03	0.25
DDB_G0293854	0	-0.22	0.15	-0.33	-0.06	-0.03	-0.02	-0.08	-1.06	-0.11
DDB_G0293954	0.84	0.06	0.05	0.19	0.01	-0.02	0.03	0.06	-1.75	0.15

Table S3

Isolate	Mating-type	*matA* - % nt identity to Ax4	MatA - % aa identity to Ax4
NC28.2	I	100	100
NC59.2	I	100	100
WS205	I	99.7(A225C)	100

Table S4

Isolate	Mating type	*matB* - % nt identity to NC66.2	MatB - % aa identity to NC66.2	*matC* - % nt identity	MatC - % aa identity	*matD* - % nt identity	MatD - % aa identity
NC42.1	II	100	100	100	100	100	100
NC94.2	II	100	100	100	100	100	100
A2cycR	II	100	100	99.8 (T360C)	100	100	100
NYA64	II	99.7	100	99.4	99	99.5	98.9
V12M2	II	100	100	100	100	100	100

Table S5

	WT-I	WT-II	WT-III	null	A	B	C	D	BC	BD	CD	BCD	S	C gen	BC gen	BCD gen
WT-I	0	70 (47)	85 (24)	0	0	0	144 (60)	0	142 (24)	0	162 (63)	246 (66)	17 (13)	107 (52)	114 (61)	337 (183)
WT-II		0	25 (21)	0	35 (13)	0	0	0	0	0	0	0	87 (48)	0	0	0
WT-III			0	0	ND	106 (41)	0	0	86 (45)	81 (22)	0	137 (127)	0	0	227 (61)	45 (24)
null				0	0	0	0	0	0	0	0	0	0	0	0	0
A					0	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
B						0	0	0	0	0	0	0	10 (4)	0	0	0
C							0	0	0	0	0	0	0	0	0	0
D								0	0	0	0	0	0	0	0	0
BC									0	0	0	0	15 (7)	0	0	0
BD										0	0	0	437 (58)	0	0	0
CD											0	0	0	0	0	0
BCD												0	545 (80)	0	0	0
S													0	0	129 (44)	275 (111)
C gen														0	0	0
BC gen															0	0
BCD gen																0

Table S6

Isolate	Mating type	*matS* (% nt identity to WS2162)	MatS (% aa identity to WS2162)	*MatT* (% nt identity to WS2162)	MatT (% aa identity to WS2162)
WS112B	III	99.8 (T67A)	99.5 (L23I)	99.9 (T1218C, T1899)	100
AC4	homothallic	89.7	77	83.3	92.2
ZA3A	homothallic	93.4	90.2	96.9	95.8

Legends to supplemental figures

Figure S1. Microarray comparative genomics provides a single candidate mating-type specific gene (full version of figure 2). Log(2)ratios comparing type II strains with our laboratory standard (type I) Ax2 were clustered using the 'average' algorithm of the R hclust function *(S5),* and the resulting dendrogram used to produce heatmaps separately for the comparisons of type I and type II strains with Ax2. Thus the order of genes is the same for both plots; again, only genes with a logratio below -2 in at least one strain are shown. Dictybase DDB_G IDs are given for each. Heatmaps were plotted using the heatmap.2 function of the gplots package *(S6)* in R.

Figure S2. The flanking genes DDB_G0289163 and DDB_G0289171 are strongly conserved. **A:** The coding sequence of DDB_G0289163 was determined from the start codon as far as the primer C9635 which was used to amplify the locus from genomic DNA (the sequence in NC66.2 is truncated because this was assembled from a *Tsp*509I-digested library and so only reaches as far as the distal 'AATT' sequence). The AX4 sequence was taken from the genome database and truncated at the same position as most of the strains sampled in the present study. Sequences were aligned using clustalw *(S37)* and plotted using texshade *(S38).* Type I, II, and III sequences are identical as far as we have sequenced. AC4 has diverged, which is consistent with previous work suggesting that it is a distant outlier within the *D. discoideum* species, or even a separate but closely related species (*39-41*). **B:** DDB_G0289171 was sequenced from its stop codon as far as the primer D2F used to amplify the locus. Sequences were aligned and plotted as above (coordinates are from the start of each sequence, not the start codon of the gene). There is more variation between strains, and at five positions, the type I and type II strains are distinguished by different variants. All of these are synonymous meaning that the amino-acid is invariant between them, and so there is no suggestion that this gene is implicated in sex-determination. Linkage of these variants with idiomorphs of the *mat* locus would not seem to be surprising given the likely close relation (in evolutionary terms and timescales) between North American isolates. The fact that NYA64 originates from Japan perhaps makes its similarity to NC66.2 more interesting. The AC4 sequence again displays most divergence, reinforcing its outlier status.

Figure S3. Global alignment of the AX4 *matA* and NC66.2 *matB* sequences. A. The nucleotide sequences of *matA* and *matB* are the same length, so texshade *(S38)* was used to highlight the identical bases (in blue). The two sequences are 65% identical to each other; this is a more than sufficient difference for our DNA microarray to distinguish between the two sequences. **B:** Alignment of MatA and MatB polypeptide sequences. Again texshade was used this time to highlight identical and related amino-acid residues between the two sequences. The polypeptides are 57% identical, and 73% similar to each other.

Figure S4. MatC and MatS are homologues, and a sequence from *Dictyostelium purpureum* is related to both. Alignments of polypeptide sequences of *Dictyostelium discoideum* (*Dd*) MatC from NC66.2 with the MatS sequence from WS2162 (type III) and their homologue from *Dictyostelium purpureum* (*Dp*) *(S42)* were made using prank *(S43)* and coloured using texshade *(S38)* in similarity mode. The *Dp* sequence was identified using blastp *(S44)* searches against a database of *Dp* predicted proteins, kindly released before publication by the Joint Genome Institute. MatC and MatS in *Dd* are about 50% identical, while the *Dp* homologue is less than 25% identical to either; the structure of the locus in *Dp* resembles that in the type III and homothallic isolates, but further investigation will be required to examine functional similarities.

Figure S5. MatD and MatT are homologues, and also related to sequences in *D. purpureum* and *Acytostelium subglobusum.* Homologues in *Dictyostelium purpureum* (*Dp*) and *Acytostelium subglobusum* (*As*) were identified by blastp *(S44).* The *As* genome database was generously made available before publication by the *Acytostelium* Genome Consortium. Alignments of polypeptide sequences of *Dictyostelium discoideum* MatD from NC66.2 with MatT sequences from WS2162 (type III), and their homologues from *Dp* and *As* were prepared as for figure S4. MatD and MatT in *Dd* are about 80% identical, the *Dp* homologue about 42% identical to these, while the *As* protein is less than 25% identical to them. Signal peptides were predicted using SignalP (*S45*): cleavage is predicted to occur after residue 19 in MatD (NC66.2 and NYA64), residue 17 in MatT (WS2162 and AC4), and after residue 18 in the *Dp* homologue. GPI anchor cleavage sites were predicted by the PredGPI software *(S46*) at positions 775 (NC66.2), 767 (WS2162 and ZA3A MatT), 763 (AC4 MatT), 755 (*Dp* homologue), and 776 (*As* homologue). All were rated as 'highly probable' except for the *Dp* homologue which was rated 'probable.' PSI-BLAST (*S44*) reveals that MatD and MatT are distantly related to the Hap2-GCS1 family *(S47*) as well as a conserved dictyostelid membrane

protein, MrhA (the *Dd* homologue is DDB_G0281923) and three *Naegleria gruberi* proteins (the N-terminal regions of XP_002675063, XP_002677524, and XP_002682216). Cysteine residues 463, 479 and 486 align with conserved cysteines in the Hap2-GCS1 domain (asterisks). The putative *As* mating-type locus is available in the EMBL database under the accession FN994780.

Figure S6. *Acytostelium subglobusum* also possesses a possible *matS* homologue. No homologue of MatC nor MatS could be identified by BLAST in the *Acytostelium subglobusum* (*As*) genome. However an open reading frame adjacent to the *matD/T* homologue in *As* is approximately the same size as the *Dictyostelium matC* and *matS* genes, and encodes a polypeptide sequence with regions of homology to MatC, using the EMBOSS *(S48*) water implementation of the Smith-Waterman algorithm (*S49*). The position of the gene next to a MatT homologue in the same divergent orientation as the MatS homologues in *Dictyostelium* supports the idea that this might be a related sequence, and given the low identity between the MatS homologues of the relatively closely related *Dictyostelium* species, it would be expected that the *As* orthologue would be even more diverged. However, experimental confirmation of a role of this gene in mating is required.
The putative *As* mating-type locus is available in the EMBL database under the accession FN994780.

Figure S1

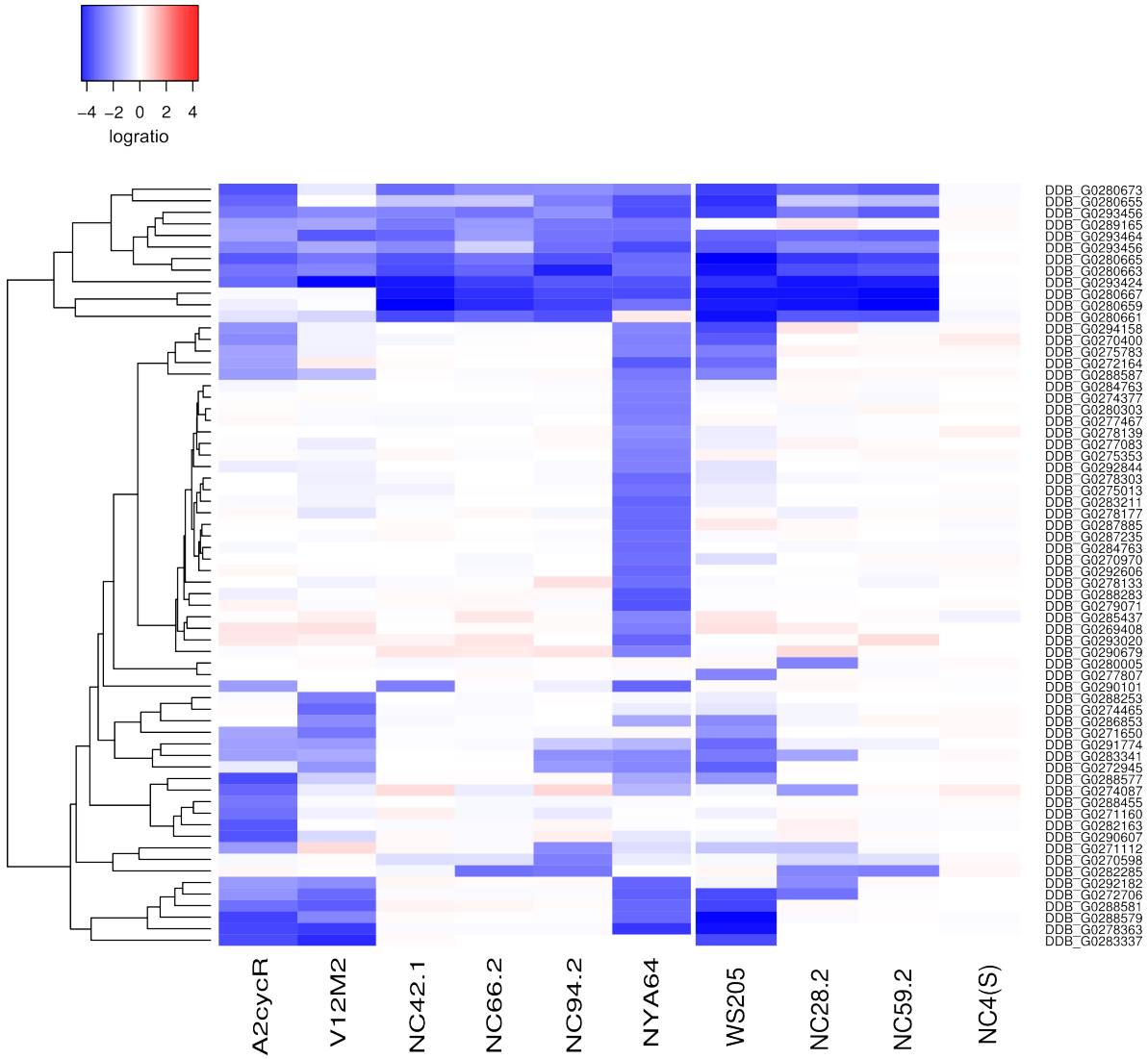
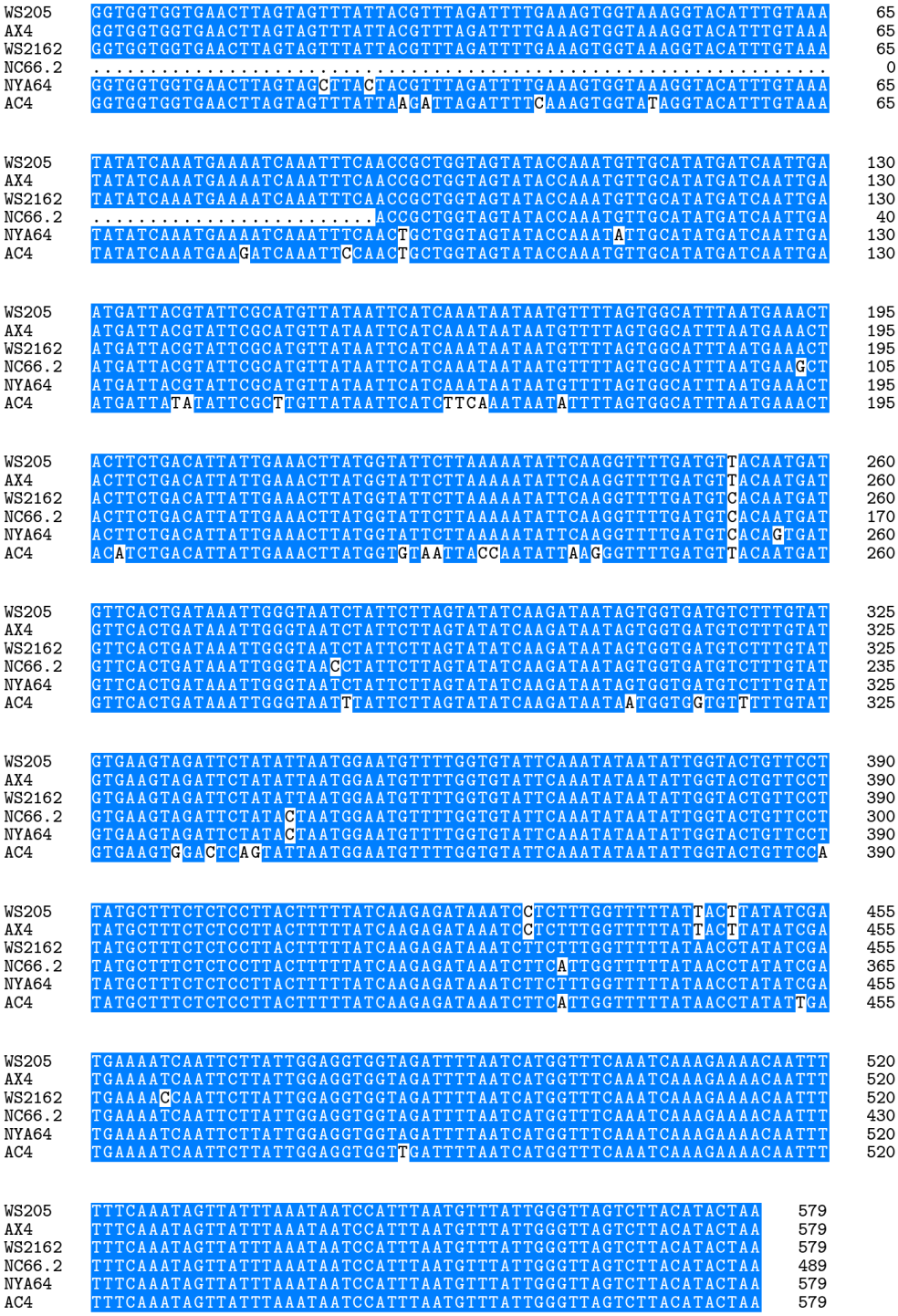

−4 −2 0 2 4
logratio
DDB_G0280673
DDB_G0280655
DDB_G0293456
DDB_G0289165
DDB_G0293464
DDB_G0293456
DDB_G0280665
DDB_G0280663
DDB_G0293424
DDB_G0280667
DDB_G0280659
DDB_G0280661
DDB_G0294158
DDB_G0270400
DDB_G0275783
DDB_G0272164
DDB_G0288587
DDB_G0284763
DDB_G0274377
DDB_G0280303
DDB_G0277467
DDB_G0278139
DDB_G0277083
DDB_G0275353
DDB_G0292844
DDB_G0278303
DDB_G0275013
DDB_G0283211
DDB_G0278177
DDB_G0287885
DDB_G0287235
DDB_G0284763
DDB_G0270970
DDB_G0292606
DDB_G0278133
DDB_G0288283
DDB_G0279071
DDB_G0285437
DDB_G0269408
DDB_G0293020
DDB_G0290679
DDB_G0280005
DDB_G0277807
DDB_G0290101
DDB_G0288253
DDB_G0274465
DDB_G0286853
DDB_G0271650
DDB_G0291774
DDB_G0283341
DDB_G0272945
DDB_G0288577
DDB_G0274087
DDB_G0288455
DDB_G0271160
DDB_G0282163
DDB_G0290607
DDB_G0271112
DDB_G0270598
DDB_G0282285
DDB_G0292182
DDB_G0272706
DDB_G0288581
DDB_G0288579
DDB_G0278363
DDB_G0283337
A2cycR
V12M2
NC42.1
NC66.2
NC94.2
NYA64
WS205
NC28.2
NC59.2
NC4(S)

Figure S2

A

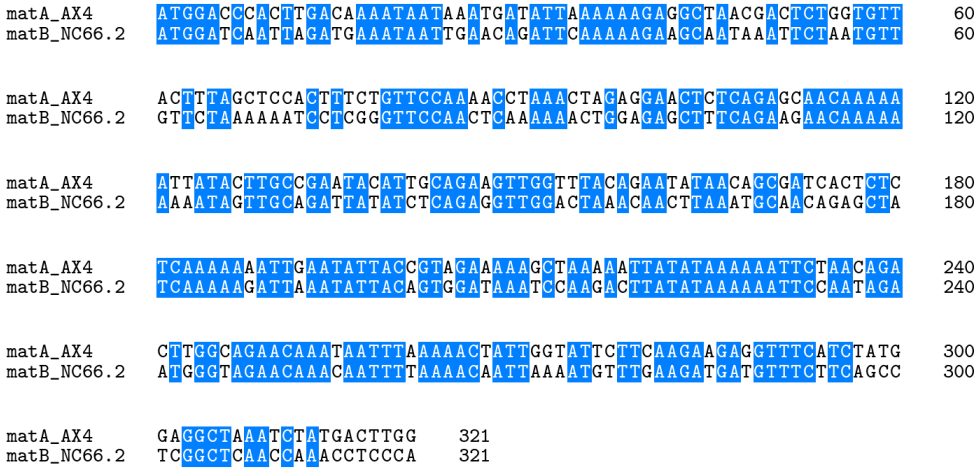

WS205 GGTGGTGGTGAACTTAGTAGTTTATTACGTTTAGATTTTGAAAGTGGTAAAGGTACATTTGTAAA 65
AX4 GGTGGTGGTGAACTTAGTAGTTTATTACGTTTAGATTTTGAAAGTGGTAAAGGTACATTTGTAAA 65
WS2162 GGTGGTGGTGAACTTAGTAGTTTATTACGTTTAGATTTTGAAAGTGGTAAAGGTACATTTGTAAA 65
NC66.2 0
NYA64 GGTGGTGGTGAACTTAGTAGCTTACTACGTTTAGATTTTGAAAGTGGTAAAGGTACATTTGTAAA 65
AC4 GGTGGTGGTGAACTTAGTAGTTTATTACATTAGATTTTCAAAGTGGTATAGGTACATTTGTAAA 65
WS205 TATATCAAATGAAAATCAAATTTCAACCGCTGGTAGTATACCAAATGTTGCATATGATCAATTGA 130
AX4 TATATCAAATGAAAATCAAATTTCAACCGCTGGTAGTATACCAAATGTTGCATATGATCAATTGA 130
WS2162 TATATCAAATGAAAATCAAATTTCAACCGCTGGTAGTATACCAAATGTTGCATATGATCAATTGA 130
NC66.2 .....ACCGCTGGTAGTATACCAAATGTTGCATATGATCAATTGA 40
NYA64 TATATCAAATGAAAATCAAATTTCAACTGCTGGTAGTATACCAAATATTGCATATGATCAATTGA 130
AC4 TATATCAAATGAAGATCAAATTCCAACTGCTGGTAGTATACCAAATGTTGCATATGATCAATTGA 130
WS205 ATGATTACGTATTCGCATGTTATAATTCATCAAATAATAATGTTTTAGTGGCATTTAATGAAACT 195
AX4 ATGATTACGTATTCGCATGTTATAATTCATCAAATAATAATGTTTTAGTGGCATTTAATGAAACT 195
WS2162 ATGATTACGTATTCGCATGTTATAATTCATCAAATAATAATGTTTTAGTGGCATTTAATGAAACT 195
NC66.2 ATGATTACGTATTCGCATGTTATAATTCATCAAATAATAATGTTTTAGTGGCATTTAATGAAGCT 105
NYA64 ATGATTACGTATTCGCATGTTATAATTCATCAAATAATAATGTTTTAGTGGCATTTAATGAAACT 195
AC4 ATGATTATATATTCGCTTGTTATAATTCATCTTCAAATAATATTTTAGTGGCATTTAATGAAACT 195
WS205 ACTTCTGACATTATTGAAACTTATGGTATTCTTAAAAATATTCAAGGTTTTGATGTTACAATGAT 260
AX4 ACTTCTGACATTATTGAAACTTATGGTATTCTTAAAAATATTCAAGGTTTTGATGTTACAATGAT 260
WS2162 ACTTCTGACATTATTGAAACTTATGGTATTCTTAAAAATATTCAAGGTTTTGATGTCACAATGAT 260
NC66.2 ACTTCTGACATTATTGAAACTTATGGTATTCTTAAAAATATTCAAGGTTTTGATGTCACAATGAT 170
NYA64 ACTTCTGACATTATTGAAACTTATGGTATTCTTAAAAATATTCAAGGTTTTGATGTCACAGTGAT 260
AC4 ACATCTGACATTATTGAAACTTATGGTGTAATTACCAATATTAAGGGTTTTGATGTTACAATGAT 260
WS205 GTTCACTGATAAATTGGGTAATCTATTCTTAGTATATCAAGATAATAGTGGTGATGTCTTTGTAT 325
AX4 GTTCACTGATAAATTGGGTAATCTATTCTTAGTATATCAAGATAATAGTGGTGATGTCTTTGTAT 325
WS2162 GTTCACTGATAAATTGGGTAATCTATTCTTAGTATATCAAGATAATAGTGGTGATGTCTTTGTAT 325
NC66.2 GTTCACTGATAAATTGGGTAACCTATTCTTAGTATATCAAGATAATAGTGGTGATGTCTTTGTAT 235
NYA64 GTTCACTGATAAATTGGGTAATCTATTCTTAGTATATCAAGATAATAGTGGTGATGTCTTTGTAT 325
AC4 GTTCACTGATAAATTGGGTAATTTATTCTTAGTATATCAAGATAATAATGGTGGTGTTTTTGTAT 325
WS205 GTGAAGTAGATTCTATATTAATGGAATGTTTTGGTGTATTCAAATATAATATTGGTACTGTTCCT 390
AX4 GTGAAGTAGATTCTATATTAATGGAATGTTTTGGTGTATTCAAATATAATATTGGTACTGTTCCT 390
WS2162 GTGAAGTAGATTCTATATTAATGGAATGTTTTGGTGTATTCAAATATAATATTGGTACTGTTCCT 390
NC66.2 GTGAAGTAGATTCTATACTAATGGAATGTTTTGGTGTATTCAAATATAATATTGGTACTGTTCCT 300
NYA64 GTGAAGTAGATTCTATACTAATGGAATGTTTTGGTGTATTCAAATATAATATTGGTACTGTTCCT 390
AC4 GTGAAGTGGACTCAGTATTAATGGAATGTTTTGGTGTATTCAAATATAATATTGGTACTGTTCCA 390
WS205 TATGCTTTCTCTCCTTACTTTTTATCAAGAGATAAATCCTCTTTGGTTTTTATTACTTATATCGA 455
AX4 TATGCTTTCTCTCCTTACTTTTTATCAAGAGATAAATCCTCTTTGGTTTTTATTACTTATATCGA 455
WS2162 TATGCTTTCTCTCCTTACTTTTTATCAAGAGATAAATCTTCTTTGGTTTTTATAACCTATATCGA 455
NC66.2 TATGCTTTCTCTCCTTACTTTTTATCAAGAGATAAATCTTCATTGGTTTTTATAACCTATATCGA 365
NYA64 TATGCTTTCTCTCCTTACTTTTTATCAAGAGATAAATCTTCTTTGGTTTTTATAACCTATATCGA 455
AC4 TATGCTTTCTCTCCTTACTTTTTATCAAGAGATAAATCTTCATTGGTTTTTATAACCTATATTGA 455
WS205 TGAAAATCAATTCTTATTGGAGGTGGTAGATTTTAATCATGGTTTCAAATCAAAGAAAACAATTT 520
AX4 TGAAAATCAATTCTTATTGGAGGTGGTAGATTTTAATCATGGTTTCAAATCAAAGAAAACAATTT 520
WS2162 TGAAAACCAATTCTTATTGGAGGTGGTAGATTTTAATCATGGTTTCAAATCAAAGAAAACAATTT 520
NC66.2 TGAAAATCAATTCTTATTGGAGGTGGTAGATTTTAATCATGGTTTCAAATCAAAGAAAACAATTT 430
NYA64 TGAAAATCAATTCTTATTGGAGGTGGTAGATTTTAATCATGGTTTCAAATCAAAGAAAACAATTT 520
AC4 TGAAAATCAATTCTTATTGGAGGTGGTTGATTTTAATCATGGTTTCAAATCAAAGAAAACAATTT 520
WS205 TTTCAAATAGTTATTTAAATAATCCATTTAATGTTTATTGGGTTAGTCTTACATACTAA 579
AX4 TTTCAAATAGTTATTTAAATAATCCATTTAATGTTTATTGGGTTAGTCTTACATACTAA 579
WS2162 TTTCAAATAGTTATTTAAATAATCCATTTAATGTTTATTGGGTTAGTCTTACATACTAA 579
NC66.2 TTTCAAATAGTTATTTAAATAATCCATTTAATGTTTATTGGGTTAGTCTTACATACTAA 489
NYA64 TTTCAAATAGTTATTTAAATAATCCATTTAATGTTTATTGGGTTAGTCTTACATACTAA 579
AC4 TTTCAAATAGTTATTTAAATAATCCATTTAATGTTTATTGGGTTAGTCTTACATACTAA 579

B


WS205 ATGGAATTTTACGAAGCCAAAGCATATTCAAATTTATTAAAAAAGAAGAAAACTGGAAATCTATT 65
AX4 ATGGAATTTTACGAAGCCAAAGCATATTCAAATTTATTAAAAAAGAAGAAAACTGGAAATCTATT 65
NYA64 ATGGAATTTTACGAAGCCAAAGCATATTCAAATTTATTAAAAAAGAAGAAAACTGGAAATCTATT 65
NC66.2 ATGGAATTTTACGAAGCCAAAGCATATTCAAATTTATTAAAAAAGAAGAAAACTGGAAATCTATT 65
WS2162 ATGGAATTTTACGAAGCCAAAGCATATTCAAATTTATTAAAAAAGAAGAAAACTGGAAATCTATT 65
AC4 ATGGAATTTTATGAAGCAAAAGCATATTCAAATTTATTAAAAAAGAAGAAAACTGGAAATCTATT 65
WS205 TAAAAATGATCCTAC...AATAACAACAACAACCACCCAAATAAATAAAGAAGAAATTATTTCTG 127
AX4 TAAAAATGATCCTAC...AATAACAACAACAACCACCCAAATAAATAAAGAAGAAATTATTTCTG 127
NYA64 TAAAAATGATCCTAC...AATAACAACAACAACCACCCAAATAAATAAAGAAGAAATTATTTCTG 127
NC66.2 TAAAAATGATCCTAC...AATAACAACAACAACCACCCAAATAAATAAAGAAGA.......... 116
WS2162 TAAAAATGATCCTAC...AATAACAACAACAACCACCCAAATAAATAAAGAAGAAATTATTTCTG 127
AC4 TAAATATGATTCTTCTTTAACAACAACAACAACCCAAATAAATAAAGAAGAAATTATTTCTG 130
WS205 ATCC 131
AX4 ATCC 131
NYA64 ATCC 131
NC66.2 116
WS2162 AT.. 129
AC4 AT.. 132

Figure S3

A


matA_AX4 ATGGACCCACTTGACAAAATAATAAATGATATTAAAAAAGAGGCTAACGACTCTGGTGTT 60
matB_NC66.2 ATGGATCAATTAGATGAAATAATTGAACAGATTCAAAAAGAAGCAATAAATTCTAATGTT 60
matA_AX4 ACTTTAGCTCCACTTTCTGTTCCAAAACCTAAACTAGAGAACTCTCAGAGCAACAAAAA 120
matB_NC66.2 GTTCTAAAAAATCCTCGGGTTCCAACTCAAAAACTGGAGAGCTTTCAGAAGAACAAAAA 120
matA_AX4 ATTATACTTGCCGAATACATTGCAGAAGTTGGTTTACAGAATATAACAGCGATCACTCTC 180
matB_NC66.2 AAAATAGTTGCAGATTATATCTCAGAGGTTGGACTAAACAACTTAAATGCAACAGAGCTA 180
matA_AX4 TCAAAAAAATTGAATATTACCGTAGAAAAAGCTAAAAATTATATAAAAAATTCTAACAGA 240
matB_NC66.2 TCAAAAAGATTAAATATTACAGTGGATAAATCCAAGACTTATATAAAAAATTCCAATAGA 240
matA_AX4 CTTGGCAGAACAAATAATTTAAAAACTATTGGTATTCTTCAAGAAGAGGTTTCATCTATG 300
matB_NC66.2 ATGGGTAGAACAAACAATTTTAAAACAATTAAAATGTTTGAAGATGATGTTTCTTCAGCC 300
matA_AX4 GAGGCTAAATCTATGACTTGG 321
matB_NC66.2 TCGGCTCAACCAAACCTCCCA 321

B

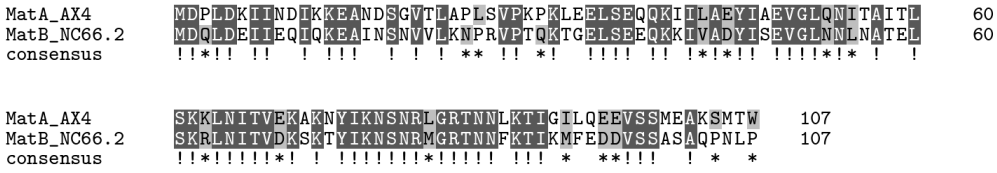

MatA_AX4 MDPLDKIINDIKKEANDSGVTLAPLSVPKPKLEELSEQQKIILAEYIAEVGLQNITAITL 60
MatB_NC66.2 MDQLDEIIEQIQKEAINSNVVLKNPRVPTQKTGELSEEQKKIVADYISEVGLNNLNATEL 60
consensus !!*!! !! ! !!! ! ! ! ** !! *! !!!! !! !*!*!! !!!!*!* ! !
MatA_AX4 SKKLNITVEKAKNYIKNSNRLGRTNNLKTIGILQEEVSSMEAKSMTW 107
MatB_NC66.2 SKRLNITVDKSKTYIKNSNRMGRTNNFKTIKMFEDDVSSASAQPNLP 107
consensus !!*!!!!!*! ! !!!!!!!*!!!!! !!! * **!!! ! * *

Figure S4

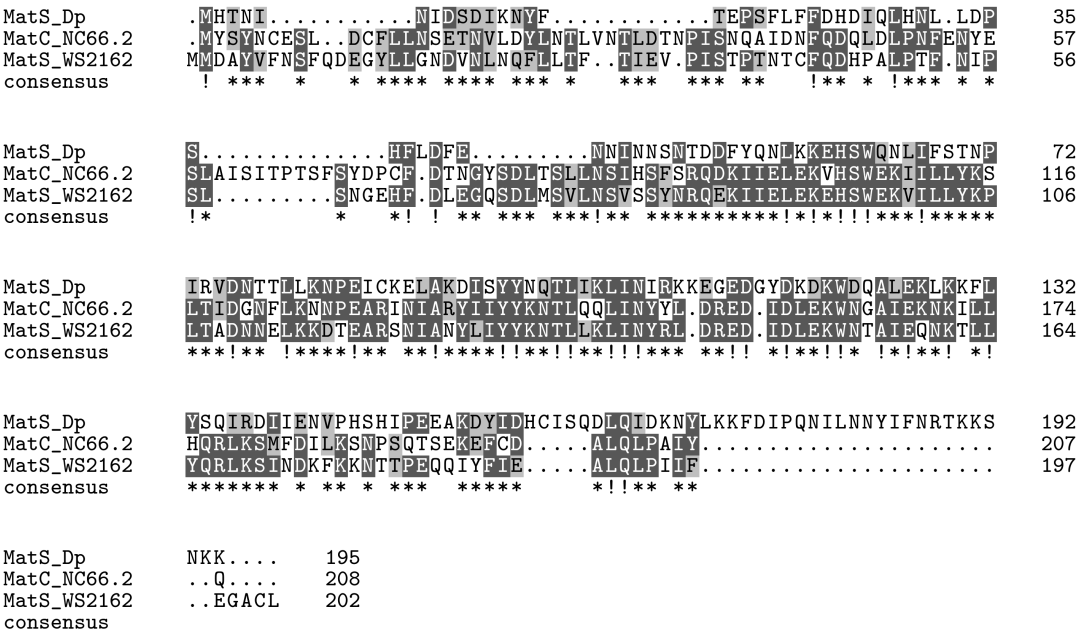
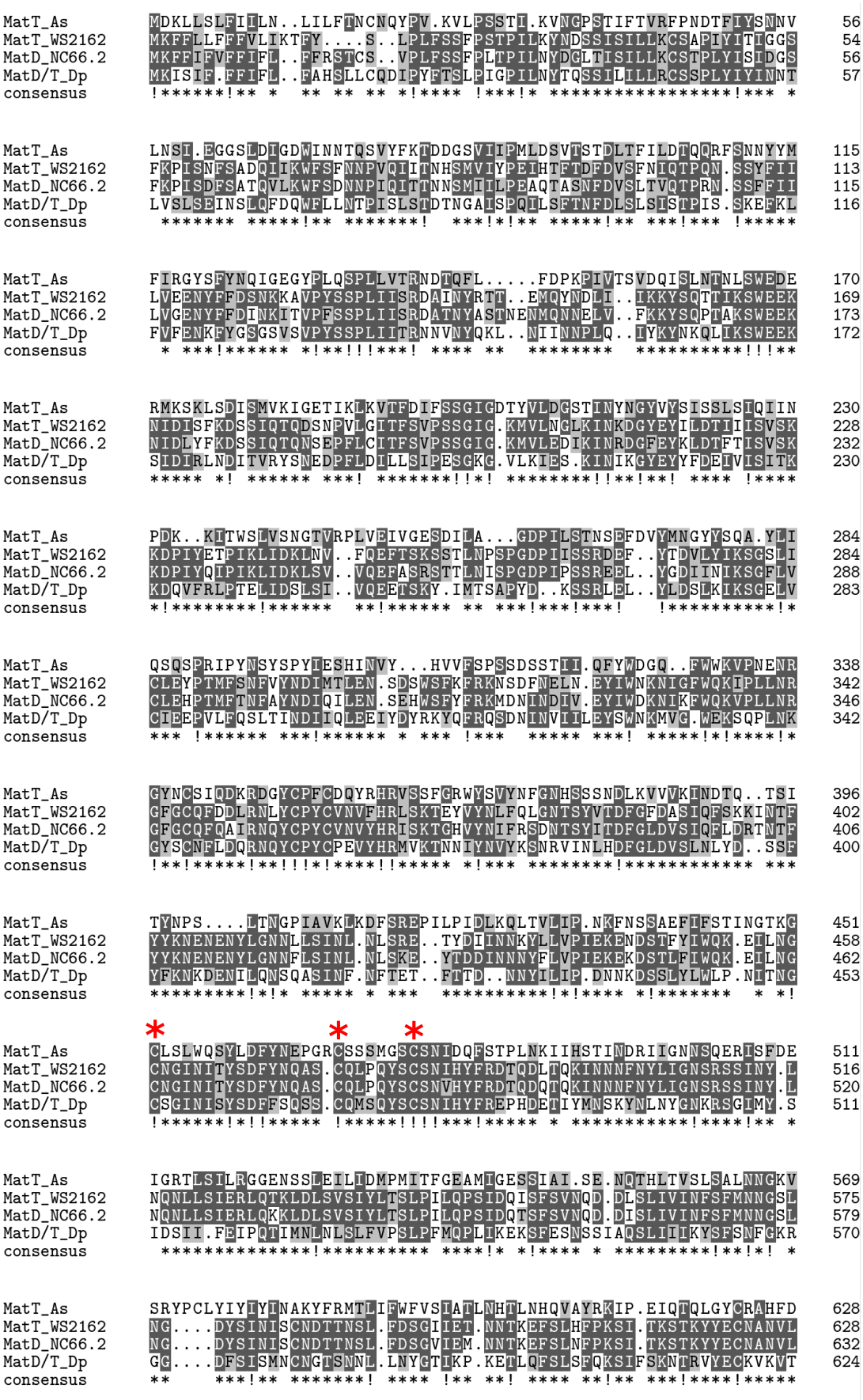

MatS_Dp .MHTNI.....NIDSDIKNYF.....TEPSFLFFDHDIQLHNL.LDP 35
MatC_NC66.2 .MYSYNCESL..DCFLLNSETNVLDYLNTLVNTLDTNPISNQAIDNFQDQLDLPNFENYE 57
MatS_WS2162 MMDAYVFNSFQDEGYLLGNDVNLNQFLLTF..TIEV.PISTPTNTCFQDHPALPTF.NIP 56
consensus ! *** * * **** **** *** * *** *** ** !** * !*** * *
MatS_Dp S.....HFLDFE.....NNINNSNTDDFYQNLKKEHSWQNLIFSTNP 72
MatC_NC66.2 SLAISITPTSFSYDPCF.DTNGYSDLTSLLNSIHSFSRQDKIIELEKVHSWEKIILLYKS 116
MatS_WS2162 SL.....SNGEHF.DLECQSDLMSVLNSVSSYNRQEKIIELEKEHSWEKVILLYKP 106
consensus !* * *! ! ** *** ***!** **********!*!*!!!***!*****
MatS_Dp IRVDNTTLLKNPEICKELAKDISYYNQTLIKLINIRKKEGEDGYDKDKWDQALEKLKKFL 132
MatC_NC66.2 LTIDGNFLKNNPEARINIARYIIYYKNTLQQLINYYL.DRED.IDLEKWNGAIEKNKILL 174
MatS_WS2162 LTADNNELKKDTEARSNIANYLIYYKNTLLKLINYRL.DRED.IDLEKWNTAIEQNKTLL 164
consensus ***!** !****!** **!****!!**!!**!!!*** **!! *!**!!* !*!**! *!
MatS_Dp YSQIRDIIENVPHSHIPEEAKDYIDHCISQDLQIDKNYLKKFDIPQNILNNYIFNRTKKS 192
MatC_NC66.2 HQRLKSMFDILKSNPSQTSEKEFCD.....ALQLPAIY..... 207
MatS_WS2162 YQRLKSINDKFKKNTTPEQQIYFIE.....ALQLPIIF..... 197
consensus ******* * ** * *** ***** *!!** **
MatS_Dp NKK.... 195
MatC_NC66.2 ..Q.... 208
MatS_WS2162 ..EGACL 202
consensus

Figure S5


MatT_As MDKLLSLFIILN..LILFTNCNQYPV.KVLPSSTI.KVNGPSTIFTVRFPNDTFIYSNNV 56
MatT_WS2162 MKFFLLFFFVLIKTFY....S..LPLFSSFPSTPILKYNDSSISILLKCSAPIYITIGGS 54
MatD_NC66.2 MKFFIFVFFIFL..FFRSTCS..VPLFSSFPLTPILNYDGLTISILLKCSTPLYISIDGS 56
MatD/T_Dp MKISIF.FFIFL..FAHSLLCQDIPYFTSLPIGPILNYTQSSILLLRCSSPLYIYINNT 57
consensus !******!** * ** ** ** *!**** !***!* ****************!*** *
MatT_As LNSI.EGGSLDIGDWINNTQSVYFKTDDGSVIIPMLDSVTSTDLTFILDTQQRFSNNYYM 115
MatT_WS2162 FKPISNFSADQIIKWFSFNNPVQIITNHSMVIYPEIHTFTDFDVSFNIQTPQN.SSYFII 113
MatD_NC66.2 FKPISDFSATQVLKWFSDNNPIQITTNNSMIILPEAQTASNFDVSLTVQTPRN.SSFFII 115
MatD/T_Dp LVSLSEINSLQFDQWFLLNTPISLSTDTNGAISPQILSFTNFDLSLSISTPIS.SKEFKL 116
consensus ******* *****!** *******! ***!*!** *****!** ***!*** !*****
MatT_As FIRGYSFYNQIGEGYPLQSPLLVTRNDTQFL.....FDPKPIVTSVDQISLNTNLSWEDE 170
MatT_WS2162 LVEENYFFDSNKKAVPYSSPLIISRDAINYRTT..EMQYNDLI..IKKYSQTTIKSWEEK 169
MatD_NC66.2 LVGENYFFDINKITVPFSSPLIISRDATNYASTNENMQNNELV..FKKYSQPTAKSWEEK 173
MatD/T_Dp FVFENKFYGSGSVSVPYSSPLIITRNNVNYQKL..NIINNPLQ..IYKYNKQLIKSWEEK 172
consensus * ***!****** *!**!!!***! **** ** ******** **********!!!**
MatT_As RMKSKLSDISMVKIGETIKLKVTFDIFSSGIGDTYVLDGSTINYNGYVYSISSLSIQIIN 230
MatT_WS2162 NIDISFKDSSIQTQDSNPVLGITFSVPSSGIG.KMVLNGLKINKDGYEYILDTIIISVSK 228
MatD_NC66.2 NIDLYFKDSSIQTQNSEPFLCITFSVPSSGIG.KMVLEDIKINRDGFEYKLDTFTISVSK 232
MatD/T_Dp SIDIRLNDITVRYSNEDPFLDILLSIPESGKG.VLKIES.KINIKGYEYYFDEIVISITK 230
consensus ***** *! ****** ***! *******!!*! ********!!**!**! **** !****
MatT_As PDK..KITWSLVSNGTVRPLVEIVGESDILA...GDPILSTNSEFDVYMNGYYSQA.YLI 284
MatT_WS2162 KDPIYETPIKLIDKLNV..FQEFTSKSSTLNPSPGDPIISSRDEF..YTDVLYIKSGSLI 284
MatD_NC66.2 KDPIYQIPIKLIDKLSV..VQEFASRSTTLNISPGDPIPSSREEL..YGDIINIKSGFLV 288
MatD/T_Dp KDQVFRLPTELIDSLSI..VQEETSKY.IMTSAPYD..KSSRLEL..YLDSLKIKSGELV 283
consensus *!********!****** **!****** ** ***!***!***!   !**********!*
MatT_As QSQSPRIPYNSYSPYIESHINVY...HVVFSPSSDSSTII.QFYWDGQ..FWWKVPNENR 338
MatT_WS2162 CLEYPTMFSNFVYNDIMTLEN.SDSWSFKFRKNSDFNELN.EYIWNKNIGFWQKIPLLNR 342
MatD_NC66.2 CLEHPTMFTNFAYNDIQILEN.SEHWSFYFRKMDNINDIV.EYIWDKNIKFWQKVPLLNR 346
MatD/T_Dp CIEEPVLFQSLTINDIIQLEEIYDYRKYQFRQSDNINVIILEYSWNKMVG.WEKSQPLNK 342
consensus *** !****** ***!****** * *** !*** ***** ***! *****!*!****!*
MatT_As GYNCSIQDKRDGYCPFCDQYRHRVSSFGRWYSVYNFGNHSSSNDLKVVVKINDTQ..TSI 396
MatT_WS2162 GFGCQFDDLRNLYCPYCVNVFHRLSKTEYVYNLFQLGNTSYVTDFGFDASIQFSKKINTF 402
MatD_NC66.2 GFGCQFQAIRNQYCPYCVNVYHRISKTGHVYNIFRSDNTSYITDFGLDVSIQFLDRTNTF 406
MatD/T_Dp GYSCNFLDQRNQYCPYCPEVYHRMVKTNNIYNVYKSNRVINLHDFGLDVSLNLYD..SSF 400
consensus !**!*****!**!!!*!****!!***** *!*** ********!************ ***
MatT_As TYNPS....LTNGPIAVKLKDFSREPILPIDLKQLTVLIP.NKFNSSAEFIFSTINGTKG 451
MatT_WS2162 YYKNENENYLGNNLLSINL.NLSRE..TYDIINNKYLLVPIEKENDSTFYIWQK.EILNG 458
MatD_NC66.2 YYKNENENYLGNNFLSINL.NLSKE..YTDDINNNYFLVPIEKEKDSTLFIWQK.EILNG 462
MatD/T_Dp YFKNKDENILQNSQASINF.NFTET..FTTD..NNYILIP.DNNKDSSLYLWLP.NITNG 453
consensus *********!*!* ***** * *** **********!*!**** *!******* * *!
MatT_As CLSLWQSYLDFYNEPGRCSSSMGSCSNIDQFSTPLNKIIHSTINDRIIGNNSQERISFDE 511
MatT_WS2162 CNGINITYSDFYNQAS.CQLPQYSCSNIHYFRDTQDLTQKINNNFNYLIGNSRSSINY.L 516
MatD_NC66.2 CNGINITYSDFYNQAS.CQLPQYSCSNVHYFRDTQDQTQKINNNFNYLIGNSRSSINY.L 520
MatD/T_Dp CSGINISYSDFFSQSS.CQMSQYSCSNIHYFREPHDEIIYMNSKYNLNYGNKRSGIMY.S 511
consensus !******!*!!***** !*****!!!!***!***** * **********!****!** *
MatT_As IGRTLSILRGGENSSLEILIDMPMITFGEAMIGESSIAI.SE.NQTHLTVSLSALNNGKV 569
MatT_WS2162 NQNLLSIERLQTKLDLSVSIYLTSLPILQPSIDQISFSVNQD.DLSLIVINFSFMNNGSL 575
MatD_NC66.2 NQNLLSIERLQKKLDLSVSIYLTSLPILQPSIDQTSFSVNQD.DISLIVINFSFMNNGSL 579
MatD/T_Dp IDSII.FEIPQTIMNLNLSLFVPSLPFMQPLIKEKSFESNSSIAQSLIIIKYSFSNFGKR 570
consensus **************!********** ****!* *!**** * *********!**!*! *
MatT_As SRYPCLYIYIYINAKYFRMTLIFWFVSIATLNHTLNHQVAYRKIP.EIQTQLGYCRAHFD 628
MatT_WS2162 NG....DYSINISCNDTTNSL.FDSGIIET.NNTKEFSLHFPKSI.TKSTKYYECNANVL 628
MatD_NC66.2 NG....DYSINISCNDTTNSL.FDSGVIEM.NNTKEFSLNFPKSI.TKSTKYYECNANVL 632
MatD/T_Dp GG....DFSISMNCNGTSNNL.LNYGTIKP.KETLQFSLSFQKSIFSKNTRVYECKVKVT 624
consensus **   ***!** *******! **** !** **! **** **!** ***!****!*****

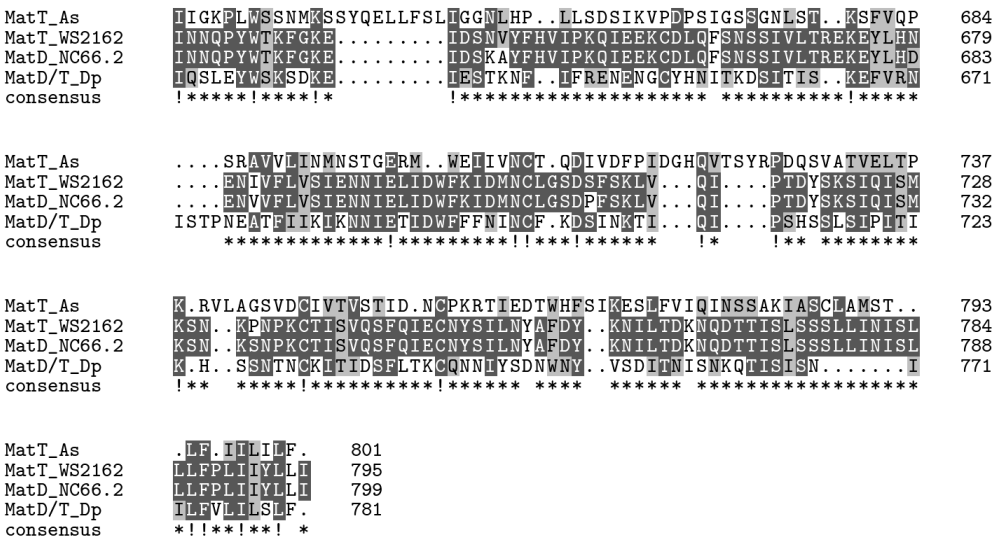
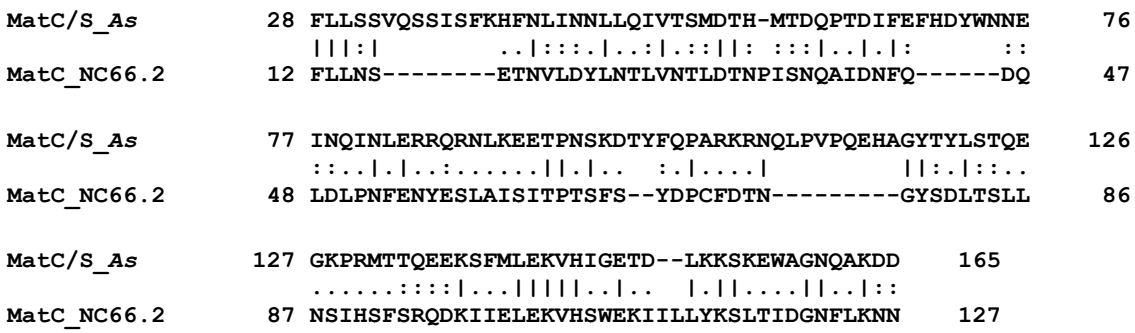

MatT_As IIGKPLWSSNMKSSYQELLFSLIGGNLHP..LLSDSIKVPDPSIGSSGNLST..KSFVQP 684
MatT_WS2162 INNQPYWTKFGKE.....IDSNVYFHVIPKQIEEKCDLQFSNSSIVLTREKEYLHN 679
MatD_NC66.2 INNQPYWTKFGKE.....IDSKAYFHVIPKQIEEKCDLQFSNSSIVLTREKEYLHD 683
MatD/T_Dp IQSLEYWSKSDKE.....IESTKNF..IFRENENGCYHNITKDSITIS..KEFVRN 671
consensus !*****!****!* !******************* **********!*****
MatT_As ....SRAVVLINMNSTGERM..WEIIVNCT.QDIVDFPIDGHQVTSYRPDQSVATVELTP 737
MatT_WS2162 ....ENIVFLVSIENNIELIDWFKIDMNCLGSDSFSKLV...QI....PTDYSKSIQISM 728
MatD_NC66.2 ....ENVVFLVSIENNIELIDWFKIDMNCLGSDPFSKLV...QI....PTDYSKSIQISM 732
MatD/T_Dp ISTPNEATFIIKIKNNIETIDWFFFNINCF.KDSINKTI...QI....PSHSSLSIPITI 723
consensus ************!*********!!***!****** !* !** ********
MatT_As K.RVLAGSVDCIVTVSTID.NCPKRTIEDTWHFSIKESLFVIQINSSAKIASCLAMST.. 793
MatT_WS2162 KSN..KPNPKCTISVQSFQIECNYSILNYAFDY..KNILTDKNQDTTISLSSSLLINISL 784
MatD_NC66.2 KSN..KSNPKCTISVQSFQIECNYSILNYAFDY..KNILTDKNQDTTISLSSSLLINISL 788
MatD/T_Dp K.H..SSNTNCKLTIDSFLTKCQNNIYSDNWNY..VSDITNISNKQTISISN.......I 771
consensus !** *****!**********!****** **** ****** ******************
MatT_As .LF.IILIIF. 801
MatT_WS2162 LLFPLIIYLLI 795
MatD_NC66.2 LLFPLIIYLLI 799
MatD/T_Dp ILFVLILSLF. 781
consensus *!!**!**! *

Figure S6


MatC/S_As 28 FLLSSVQSSISFKHFNLINNLLQIVTSMDTH-MTDQPTDIFEFHDYWNNE 76
MatC_NC66.2 12 FLLNS--------ETNVLDYLNTLVNTLDTNPISNQAIDNFQ------DQ 47
MatC/S_As 77 INQINLERRQRNLKEETPNSKDTYFQPARKRNQLPVPQEHAGYTYLSTQE 126
MatC_NC66.2 48 LDLPNFENYESLAISITPTSFS--YDPCFDTN---------GYSDLTSLL 86
MatC/S_As 127 GKPRMTTQEEKSFMLEKVHIGETD--LKKSKEWAGNQAKDD 165
MatC_NC66.2 87 NSIHSFSRQDKIIELEKVHSWEKIILLYKSLTIDGNFLKNN 127

Supplemental references

S1. M. Sussman, *Methods Cell Biol.* **28**, 9-29 (1987).

S2. G. Bloomfield, Y. Tanaka, J. Skelton, A. Ivens, R. R. Kay, *Genome Biol.* **9**, R75 (2008).

S3. M. Berks, R. R. Kay, *Development* **110**, 977-984 (1990).

S4. G. Smyth, in: Bioinformatics and Computational Biology Solutions Using R and Bioconductor. R. Gentleman, V. Carey, W. Huber, R. Irizarry, S. Dudoit, Eds. (Springer-Verlag New York Inc., 2005), pp. 397-420.

S5. R Development Core Team, R: A language and environment for statistical computing, http://www.R-project.org/

S6. G. Warnes, includes R source code and/or documentation contributed by Ben Bolker and Thomas Lumley; gplots: Various R programming tools for plotting data.

S7. J. Faix, L. Kreppel, G. Shaulsky, M. Schleicher, A. R. Kimmel, *Nucleic Acids Res.* **32**, e143 (2004).

S8. K. M. Pang, M. A. Lynes, D. A. Knecht, *Plasmid* **41**, 187-197 (1999).

S9. A. Barth, A. Müller-Taubenberger, P. Taranto, G. Gerisch, *J. Cell Biol.* **124**, 205-215 (1994).

S10. M. Fischer, I. Haase, E. Simmeth, G. Gerisch, A. Müller-Taubenberger, *FEBS Let.t* **577**, 227-232 (2004).

S11. M. L. W. Knetsch, G. Tsiavaliaris, S. Zimmermann, U. Rühl, D. J. Manstein, *J. Muscle Res. Cell. Motil.* **23**, 605-611 (2002).

S12. D. M. Veltman, G. Akar, L. Bosgraaf, P. J. Van Haastert, *Plasmid* **61**, 110-118 (2009).

S13. M. A. Clark, D. Francis, R. Eisenberg, *Biochem. Biophys. Res. Commun.* **52**, 672-678 (1973).

S14. G. W. Erdos, K. B. Raper, L. K. Vogen, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 1828-1830 (1973).

S15. H. Urushihara, T. Muramoto, *Eur. J. Cell Biol.* **85**, 961-968 (2006).

S16. G. E. Robson, K. L. Williams, *Genetics* **93**, 861-875 (1979).

S17. G. E. Robson, K. L. Williams, *Curr. Genet.* **1**, 229-232 (1980).

S18. M. Demerec, E. A. Adelberg, A. J. Clark, P. E. Hartman, *Genetics* **54**, 61-76 (1966).

S19. (available at http://www.dictybase.org/Dicty_Info/nomenclature_guidelines.html).

S20. R. L. Metzenberg, N. L. Glass, *BioEssays* **12**, 53-59 (1990).

S21. L. Eichinger et al., *Nature* **435**, 43-57 (2005).

S22. M. A. Wallace, K. B. Raper, *J Gen. Microbiol.* **113**, 327-337 (1979).

S23. D. L. Welker, K. L. Williams, *Genetics* **102**, 691–710 (1982).

S24. E. R. Katz, M. Sussman, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 495–498 (1972).

S25. D. Mosses, K. L. Williams, P. C. Newell, *J. Gen. Microbiol.* **90**, 247-259 (1975).

S26. C. Miceli, A. La Terza, R. A. Bradshaw, P. Luporini, *Eur. J. Biochem.* **202**, 759-764 (1991)

S27. L. A. Casselton, *Heredity* **88**, 142-147 (2002).

S28. J. Kämper, M. Reichmann, T. Romeis, M. Bölker, R. Kahmann, *Cell* **81**, 73-83 (1995).

S29. A. C. Rea, J. B. Nasrallah, *Int. J. Dev. Biol.* **52**, 627-636 (2008).

S30. R. H. Schwartz, *Annu. Rev. Immunol.* **3**, 237-261 (1985)

S31. A. W. Nickerson, K. B. Raper, *Am. J. Bot.* **60**, 190-197 (1973).

S32. D. Francis, *Genetics* **148**, 1829-1832 (1998).

S33. K. B. Raper, *J. Agr. Res.* **50**, 135-147 (1935).

S34. D. Francis, R. Eisenberg, *Mol. Ecol.* **2**, 385-391 (1993).

S35. C. Kanz et al., *Nucleic Acids Res.* **33**, D29-33 (2005).

S36. P. Fey et al., *Nucleic Acids Res.* **37**, D515-519 (2009).

S37. M. Larkin et al., *Bioinformatics* **23**, 2947-2948 (2007).

S38. E. Beitz, *Bioinformatics* **16**, 135-139 (2000).

S39. D. A. Briscoe, A. A. Gooley, R. L. Bernstein, G. M. McKay, K. L. Williams, *Genetics* **117**, 213-220 (1987).

S40. W. B. Evans, J. E. Hughes, D. L. Welker, *Genetics* **119**, 561–569 (1988).

S41. P. Schaap et al., *Science* **314**, 661-663 (2006).

S42. (available at http://genome.jgi-psf.org/Dicpu1/Dicpu1.home.html).

S43. A. Loytynoja, N. Goldman, *Science* **320**, 1632-1635 (2008).

S44. S. F. Altschul et al., *Nucleic Acids Res.* **25**, 3389-3402 (1997).

S45. J. D. Bendtsen, H. Nielsen, G. von Heijne, S. Brunak, *J. Mol. Biol.* **340**, 783-795 (2004).

S46. A. Pierleoni, P. L. Martelli, R. Casadio, *BMC Bioinformatics* **9**, 392 (2008).

S47. J. L. Wong, M. A. Johnson, *Trends Cell Biol.* **20**, 134-141 (2010).

S48. P. Rice, I. Longden, A. Bleasby, *Trends in Genetics* **16**, 276-277 (2000).

S49. T. F. Smith, M. S. Waterman, *J. Mol. Biol.* **147**, 195-197 (1981).