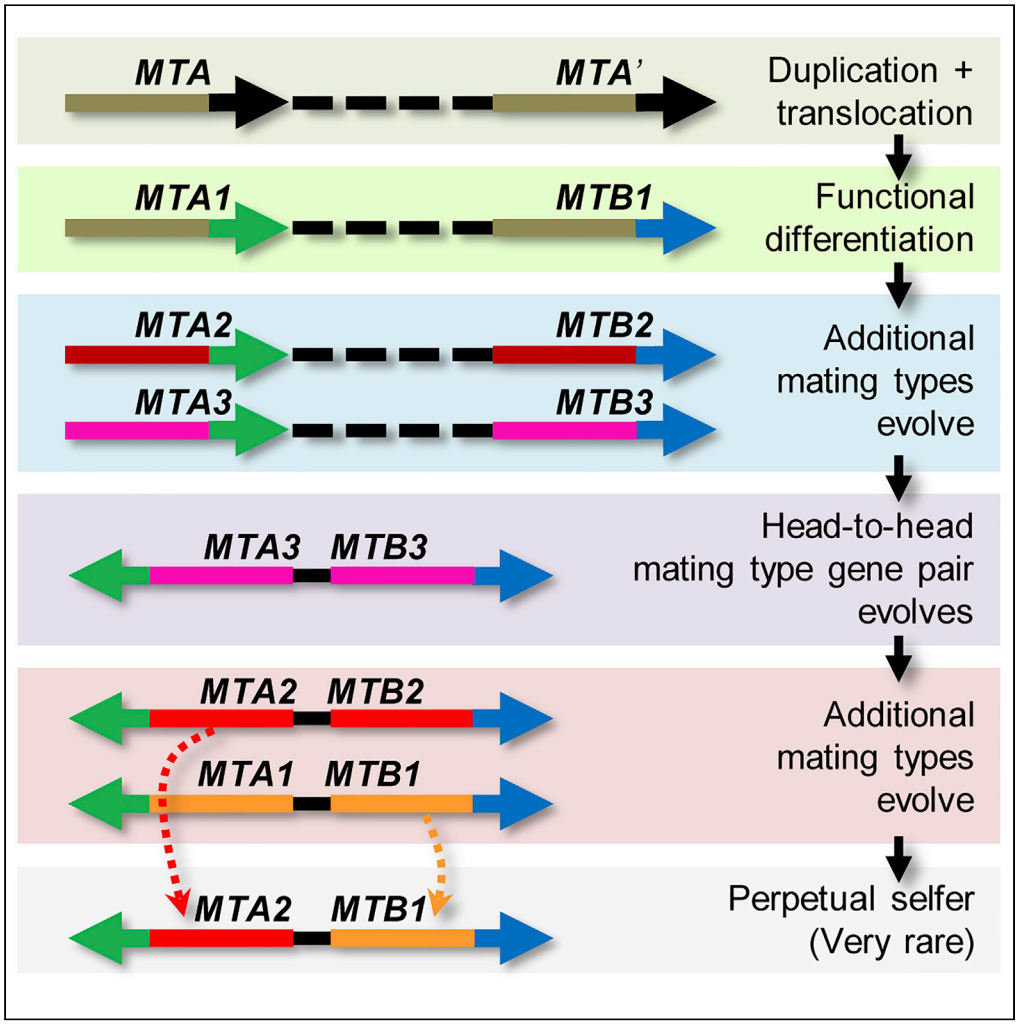
iScience

CellPress
OPEN ACCESS

Article

Evolution of the mating type gene pair and multiple sexes in *Tetrahymena*



Guanxiong Yan, Wentao Yang, Xiaojie Han, ..., Eileen P. Hamilton, Eduardo Orias, Wei Miao

orias@lifesci.ucsb.edu (E.O.)
miaowei@ihb.ac.cn (W.M.)

Highlights

The two *Tetrahymena* mating type proteins evolved from a common ancestor

Successive replacement waves generated the current diversity of mating type proteins

Well defined segments of both mating type proteins show differential lineage sorting

Perpetual selfer, *T. shanghaiensis*, has a heterospecific mating type gene pair

Yan et al., iScience *24*, 101950
January 22, 2021
https://doi.org/10.1016/j.isci.2020.101950

iScience

CellPress
OPEN ACCESS

Article

Evolution of the mating type gene pair and multiple sexes in *Tetrahymena*

Guanxiong Yan,[1,3,4] Wentao Yang,[1,3,4] Xiaojie Han,[4,6] Kai Chen,[1,4] Jie Xiong,[1,4] Eileen P. Hamilton,[2] Eduardo Orias,[2,*] and Wei Miao[1,3,4,5,7,*]

Summary

The multiple mating type system of the Ciliate *Tetrahymena thermophila* is a self/non-self recognition system, whose specificity resides in a head-to-head, functionally distinct pair of genes, *MTA and MTB*. We have now sequenced and analyzed these mating type genes in nine additional *Tetrahymena* species. We conclude that *MTA* and *MTB* are derived from a common ancestral gene and have co-evolved for at least ~150 Myr. We show that *T. shanghaiensis*, a perpetual selfer (unisexual) species, has a single mating type gene pair, whose *MTA* and *MTB* genes likely have different mating type specificity. We document the recent replacement of a complete different set of mating type specificities for another, illustrating how quickly this can happen. We discuss how varying conditions of reproductive stress could result in evolutionary co-adaptations of *MTA* and *MTB* genes and changes in mating type determination mechanisms.

Introduction

Sex is an evolutionary conserved process among organisms, including the Ciliated Protozoa (= Ciliates), a unicellular eukaryotic phylum (Bachtrog et al., 2014). Mating type systems generally ensure sexual self-incompatibility and promote outbreeding. Most eukaryotic species have two mating types or sexes and thus a binary mating system. However, systems with more than two mating types exist in some groups such as ciliates and mushrooms (Phadke and Zufall, 2010; Kües, 2015). Ciliate mating systems are very diverse: they vary in such features as number of mating types, mechanism of mating type determination (MTD), and molecular nature of the mating type proteins. This suggests that many fundamental changes in mating type biology have independently evolved in the major Ciliate clades (Phadke and Zufall, 2010). The *Tetrahymena* genus of ciliates is thus an excellent model system for studying multiple mating type systems and their evolution.

As in other ciliates, cells of most *Tetrahymena* species possess two kinds of nuclei: a diploid, silent germline nucleus (the micronucleus or MIC) and a polyploid, highly expressed somatic nucleus (the macronucleus or MAC). The *Tetrahymena* life cycle consists of two stages: asexual reproduction by binary fission when food is abundant and conjugation triggered by starvation (reviewed in (Orias et al., 2011; Orias et al., 2017)). Key life cycle features of genetic significance, illustrated in Figure 1, are:

- Only one of the four MIC meiotic products is retained in each conjugant.
- Reciprocal fertilization generates genetically identical, diploid zygote nuclei in each conjugant.
- The zygote nucleus in each conjugant divides twice mitotically, two products are the new MICs, while the other two differentiated into the new MACs.
- During MAC differentiation the five MIC chromosomes undergo programmed site-specific fragmentation, resulting in 180 MAC chromosomes. These acentromeric chromosomes are then amplified (~45 G1 copies).
- The two exconjugant cells from a pair divide, resulting in four cells with genetically identical MICs but independently differentiated MACs, called "karyonides".

[1]State Key Laboratory of Freshwater Ecology and Biotechnology, Wuhan 430072, China

[2]Department of Molecular, Cellular and Developmental Biology, University of California Santa Barbara, Santa Barbara, CA 93106, USA

[3]University of Chinese Academy of Sciences, Beijing 100049, China

[4]Key Laboratory of Aquatic Biodiversity and Conservation, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China

[5]CAS Center for Excellence in Animal Evolution and Genetics, Kunming 650223, China

[6]College of Fisheries and Life Science, Dalian Ocean University, Dalian 116023, China

[7]Lead contact

*Correspondence: orias@lifesci.ucsb.edu (E.O.), miaowei@ihb.ac.cn (W.M.)

https://doi.org/10.1016/j.isci.2020.101950

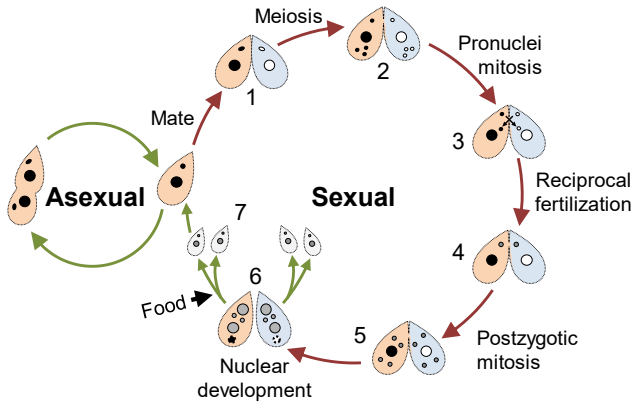


Figure 1. Schematic diagram of the *Tetrahymena* life cycle

The life cycle of sexual *Tetrahymena* species consists of two stages: sexual and asexual. To start the sexual stage, two starved cells conjugate (1). The micronucleus (MIC) of each conjugant first undergoes meiosis (2). A randomly chosen meiotic product in each conjugant divides mitotically and generates two gamete pronuclei (3). Next, the two conjugants exchange one of their two gamete pronuclei, after which the pronuclei fuse to form a genetically identical diploid fertilization nucleus in each conjugant (4); thus, each conjugant gets a haploid genome from each parent. The fertilization nucleus undergoes two rounds of mitotic division, generating four nuclei in each conjugant (5). Two of these nuclei differentiate into new macronuclei (MACs) and the other two remain MICs (6). The old MAC is degraded, and the two "exconjugants" separate from one another. After feeding, the exconjugants undergo their first postzygotic fission; each of the new MACs is distributed into a different daughter cell, the "karyonide" cells (7). At this stage, the one MIC and one MAC condition has been restored. Subsequently, during the asexual part of the life cycle, the MIC divides mitotically and the MAC divides amitotically at every succeeding cell cycle. Amitotic division of the polyploid MAC results in the random segregation of parental chromosomes among daughter cells, leading to segregation of genetic diversity among individuals (allelic assortment). Ultimately, by asexual reproduction and allelic assortment, an individual MAC tends to become homozygous for its entire genome.

- Phenotypic assortment: when asexual multiplication resumes after conjugation, amitotic division of the polyploid MAC results in the random distribution of daughter chromosome copies at every fission. This allows segregation at all heterozygous loci present in a newly differentiated MAC and ultimately generates whole-genome homozygous MACs.
- Sexual progeny are initially sexually immature; in *T. thermophila* they must undergo 40–70 fissions before they reach sexual maturity and can mate again.

The diversity of *Tetrahymena* mating type systems was described in detail in a series of papers by Nanney, Elliott, and their collaborators, beginning in the early 1950s (reviewed in (Orias, 1981; Orias et al., 2017)). Among sexual species in this genus, the number of known mating types per species ranges from three to nine. Certain rare species (e.g., *T. shanghaiensis*) are unisexual (selfers), meaning that sexually mature cells in a clonal population mate with one another upon starvation (Chen et al., 1982; Simon et al., 2009) and all their progeny are selfers. Cells of other species (e.g., *T. pyriformis* and *T. vorax*) have lost their MIC (i.e. the germline nucleus) and thus can only reproduce asexually (Gruchy, 1955; Doerder, 2014).

The best characterized species belong to either the "Australis" or the "Borealis" clades, which diverged ~150 Myr ago (Xiong et al., 2019). Earlier studies suggested that *Tetrahymena* MTD patterns can also be classified into two categories that co-branch with the phylogenetic tree. Investigated species in the "Australis" clade exhibit "synclonal MTD"; where the four genetically identical karyonides of a mating pair (the synclone) express the same mating type with a Mendelian inheritance pattern when they reach sexual maturity (Figure S1A). In *T. pigmentosa*, for example, mating types are controlled by three alleles of a single *mat* locus that show "peck-order" dominance (Simon, 1980). In contrast, investigated species in the "Borealis" clade show "karyonidal MTD", where mating type is randomly and independently determined in each new MAC. This results in four genetically identical karyonides which often express different mating types (Figure S1B).

The molecular basis of mating type specificity has only been investigated in *T. thermophila* ("Borealis" clade). Mating type is determined by a mating type gene pair (mtGP), a head-to-head arrangement of

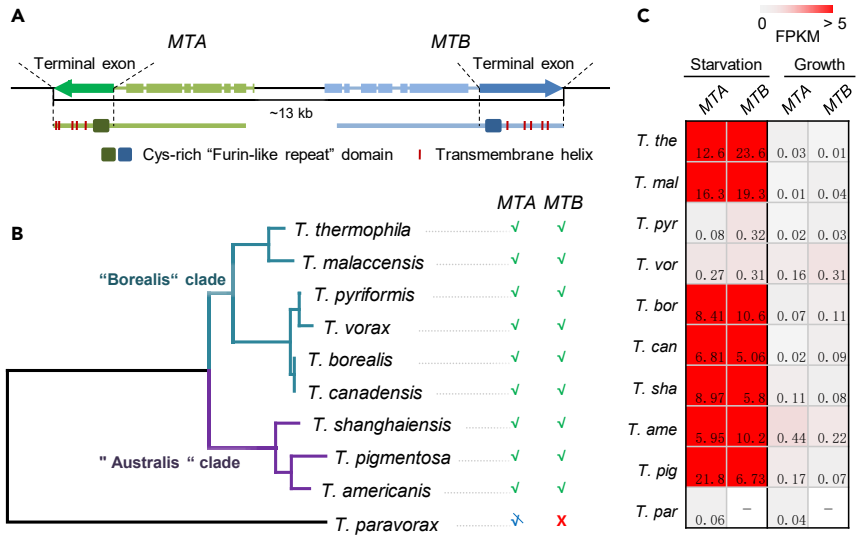


Figure 2. *Tetrahymena* mating type genes, species phylogeny and mating type gene expression

(A) Schematic diagram of a *T. thermophila* mtGP and the domain composition of the two encoded proteins. Top line: *MTA* gene (green), *MTB* gene (blue), and intergenic region (black). Blue vs. green colors are used to indicate that, other than the furin-like repeats, there is no sequence conservation between MTA and MTB proteins. Darker color thick lines: conserved 3′ terminal exons; lighter color thick lines: mating type-specific exons. Bottom line: MTA and MTB proteins, showing detail of terminal exon domains conserved in all *Tetrahymena* mating type genes. Vertical red line: predicted transmembrane helices; squares: cysteine-rich "Furin-like repeat" domain. The *MTA* 3′ terminal exon sequences are >99% conserved among six published mating types, and likewise for the *MTB* genes.

(B) *Tetrahymena* species phylogeny and *MTA*, *MTB* gene homologs present in each species. The phylogenetic tree is based on 18S rRNA sequences. Green check mark, homologous gene identified; blue strikeout check mark, homologous gene with low sequence similarity identified; red ×, no homologous gene identified.

(C) Gene expression measurements, using RNA sequencing data, for mating type genes of all ten species. Numbers represent FPKM values. All mating types were measured and averages are shown when data was collected for more than one mating type in the species.

two mating type genes (*MTA* and *MTB*) (Cervantes et al., 2013) (Figure 2A). In this study, we investigated the molecular evolution of *Tetrahymena* mtGPs in a phylogenetically wide range of *Tetrahymena* species, including two asexual species (*pyriformis* and *vorax*), a unisexual obligatory selfer species (*shanghaiensis*), and the species furthest removed from the "Australis" and "Borealis" clades (*paravorax*) (Figure 2B). We provide evidence for the evolution of all *Tetrahymena* mating type proteins from an ancient member of the "Furin-like repeat" protein family, for the coevolution of *MTA* and *MTB* genes, for the evolution of a heterotypic *MTA-MTB* gene pair leading to perpetual selfing, and for the recent replacement of one multiple mating type system with another within a subgroup of the genus *Tetrahymena*.

Results

Mating type gene pair homologs exist in all but the most distantly related *Tetrahymena* species examined

In *T. thermophila*, the *MTA* and *MTB* genes shares some similar features but have completely different sequences. The sequence of the terminal exons is highly conserved between *MTA* genes encoding different mating types (Figure 2A) (nucleotide identity >0.99); the same is true among the *MTB* alleles. In contrast, the remainder of the *MTA* and *MTB* genes and the intergenic region are mating type-specific (nucleotide identity <0.6). The 3′-terminal exons of the *MTA* and *MTB* genes both encode five predicted transmembrane helices and a cysteine-rich "Furin-like repeat" domain. These features are diagrammed in Figure 2A; see also (Cervantes et al., 2013; Orias et al., 2017).

We looked for homologs of the *T. thermophila* MAC mating type genes in nine additional species spanning the ~300 Myr old *Tetrahymena* genus (Xiong et al., 2019) (Table 1). Six species (*thermophila*, *malaccensis*, *pyriformis*, *vorax*, *borealis* and *canadensis*) are in the "Borealis" clade, three (*shanghaiensis*, *americanis*, and *pigmentosa*) are in the "Australis" clade, and one (*paravorax*) diverged from both clades at the

Table 1. Mating type systems of *Tetrahymena* species investigated in this article

Clade[a]	Subclade[a]	Species	# Mating types	MTD pattern[b]
Borealis	"The-Mal"	*Thermophila*	7	Karyonidal
Borealis	"The-Mal"	*Malaccensis*	6	Karyonidal
Borealis	"Pyr-Vor"	*Pyriformis*	Asexual	NA
Borealis	"Pyr-Vor"	*Vorax*	Asexual	NA
Borealis	"Bor-Can"	*Borealis*	7	Synclonal
Borealis	"Bor-Can"	*Canadensis*	5	Synclonal
Australis	"Pig-Ame"	*shanghaiensis*	Perpetual selfer	NA,
Australis	"Pig-Ame"	*pigmentosa*	3	Synclonal
Australis	"Pig-Ame"	*americanis*	9	Synclonal
Paravorax	NA	*paravorax*	ND	ND

N/D: not determined.

[a]The phylogeny is illustrated in Figure 2B.

[b]Mating type determination pattern observed in sexual progeny (see text and Figure S1 for explanation). N/A: not applicable; only strain and only species characterized in this clade.

base of the *Tetrahymena* genus (Figure 2B). *T. pyriformis* and *vorax* cells never mate, they lack an MIC and only reproduce asexually. And to our knowledge, sexual reproduction has not been observed in *paravorax*. Altogether, we investigated mating type genes in 19 *Tetrahymena* strains (Table 2). We found mating type gene homologs in all sexual species, as well as in three asexual species (*pyriformis*, *vorax* and *paravorax*) and in the unisexual strain, *shanghaiensis* (a "selfer" species, in which starvation of sexual mature cells triggers intraclonal mating). Within each sexual species, we verified that mating only occurs between starved cells of different mating types; no mating was observed between starved cells from different species (for experimental details see Transparent methods).

Our searches revealed a single, mating type-specific mtGP with *MTA* and *MTB* homologs in head-to-head orientation for each mating type of every species, with the sole exception of the *T. paravorax* strain, which has a truncated *MTA* gene (named *MTAL* for *MTA*-like) and lacks an *MTB* homolog (Figures 2B and S2). Additionally, we determined that all mtGPs of species with sequenced genomes (Xiong et al., 2019), with the single exception of the asexual *T. vorax* mtGP, have syntenic chromosomal locations (Figure S3).

It has long been known that starvation conditions are required for conjugation in *Tetrahymena*. Consistent with this, the *T. thermophila MTA* and *MTB* genes are highly expressed during starvation, but are essentially silent during vegetative growth (Cervantes et al., 2013). To investigate whether mtGPs of other *Tetrahymena* species might function similarly in mating, their expression levels were measured during growth and starvation (Figure 2C). In the seven sexual species (including the "selfer" species, *T. shanghaiensis*), the expression pattern of the mating type genes is identical to that of *T. thermophila*, consistent with conservation of mtGP function in mating. In the asexual species (*pyriformis*, *vorax*, and *paravorax*), the mating type genes are not induced by starvation, suggesting that they are no longer functional in mating.

Tetrahymena MTA and *MTB* genes are likely derived from a common ancestral gene

All 37 *MTA* and *MTB* genes reported in the previous section belong to the same gene superfamily which has a conserved cysteine-rich "Furin-like repeat" domain (Figures 3A and S2). Alignment of the "Furin-like repeat" domains in all MTA and MTB homologs shows that all 14 cysteine residues are highly conserved with one another, although four cysteines are missing in MTAL (Figure 3A). Since cysteine residues play many important roles, such as forming covalent disulfide bonds with each other (Sela and Lifson, 1959; Thornton, 1981), the conservation of these residues may be essential to the secondary structure of the mating type proteins and their function.

The full length *Tetrahymena MTA* and *MTB* genes investigated here contain 5-8 introns. In the sexual species, four introns (#3, #5, #6, and #7) show total conservation of phase and approximate location in the 32 sequenced *MTA* and *MTB* genes (Figure 3B and Table S1, details in Figures S2 and S4). Intron #7 precedes the 3'-terminal exon, which encodes the "Furin-like repeat" domain and transmembrane helices of the

Table 2. Relevant strain and mating type information on the *Tetrahymena* strains used in this work

Species	Strain ID[a]	Mating type[b]
thermophila	SD01580	II
thermophila	SD01653	III
thermophila	SD01582	IV
thermophila	SD01656	V
thermophila	SD01584	VI
thermophila	SD01585	VII
Malaccensis	SD01608	X
Pyriformis	SD00707	NA
Vorax	SD30421	NA
borealis	SD01609	X
borealis	SD19502	Y
borealis	SD19803	Z
canadensis	SD30770	X
shanghaiensis	SD205039	Selfer
pigmentosa	SD19481	III
pigmentosa	SD20427	I
americanis	SD21194	X
americanis	SD21244	Y
paravorax	SD205177	N/D

N/A: not applicable; these strains lack a micronucleus and are asexual. N/D: not determined.
[a]Tetrahymena Stock Center ID numbers.
[b]Where the relationship to previously published mating types is undetermined, we have used capital letters X, Y, and Z to avoid confusing the literature.

MTA, *MTAL*, and *MTB* genes. A fifth intron (#2) is conserved among the *MTA* and *MTB* genes of all mtGPs, with the exception of *T. canadensis*, which has just one sequenced mtGP (*MTAX*).

The conservation of intron location, in both *MTA* and *MTB*, is particularly striking within subgroup alignments (Figure S4) that include at least two species for each subgroup. Astonishingly, 27 (of 38) introns occur at identical codon locations within their gene. In another seven cases, the introns are located at an adjacent codon. In the four remaining cases, the intron is located a few codons further away, but there are clearly insertions/deletions of one or more codons in the immediate neighborhood. The only outlier is intron #2 in the "Bor-Can" *MTA*s, which has changed phase in *T. borealis MTAZ* and is missing in *T. canadensis MTAX*.

The absolute conservation of the "Furin-like repeat" domain, together with the high degree of conservation of intron location and phase in such a large (~1,500 aa) protein, along with chromosomal synteny, are strong evidence that all *MTA* and *MTB* genes were derived from a common ancestral gene which existed prior to the divergence of the "Australis" and "Borealis" clades in *Tetrahymena*, ~150 Myr ago (Xiong et al., 2019).

To confirm the evolutionary relationship between MTA and MTB proteins, we generated phylogenetic trees of these proteins from the sexual species. For each protein, we made separate trees for the entire protein (Figure S5), for the C-terminal, transmembrane exon (distal ~1/3 of the protein), and for the rest of the protein (proximal ~2/3 of the protein (Figures 4A and 4B, respectively)). The results show that all three trees have two main branches, such that all MTA proteins fall cleanly into one branch, while all MTB proteins fall cleanly into the other branch. This is consistent with the nearly complete lack of overall sequence similarity observed between the MTA and MTB proteins when they are co-aligned (Figure S2). Thus the trees support the conclusion that the *MTA* and *MTB* genes diverged structurally and functionally from a common ancestral gene early in *Tetrahymena* evolution, prior to the divergence of the "Australis" and "Borealis" clades.

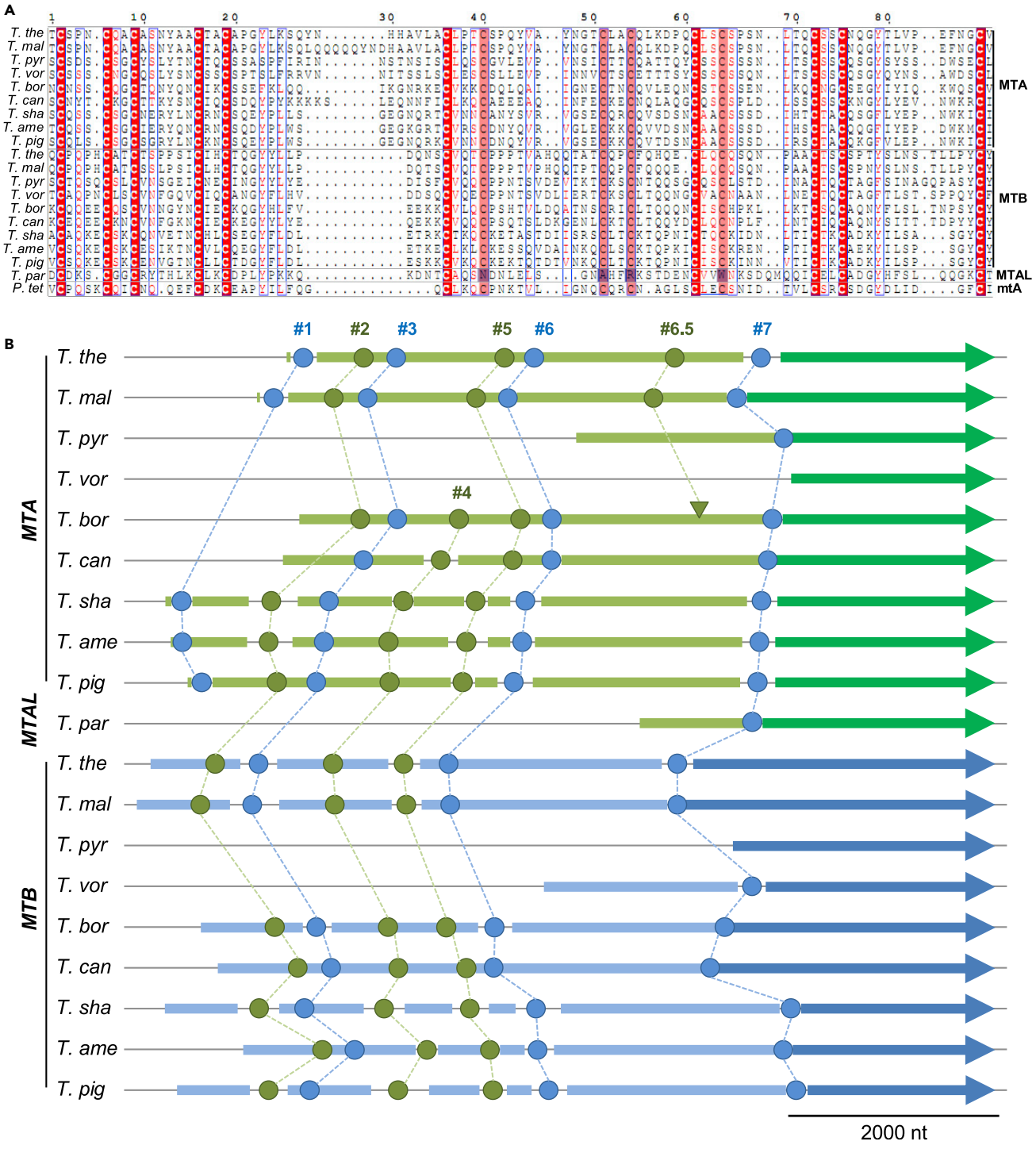


Figure 3. Highly conserved features of *Tetrahymena MTA* and *MTB* genes: FLR repeats and introns

(A) Sequence alignment of the "Furin-like repeat" domains. Cysteine sites are conserved in MTA and MTB proteins; note that four conserved cysteines (darker background) are missing in *T. paravorax* MTAL. Pink letters: partially conserved amino acids. P.tet: *Paramecium tetraurelia* mtA gene.

(B) Conservation of intron location and phase among *MTA* and *MTB* genes. Thick green lines, *MTA* exons; thick blue lines, MTB exons; thick dark green or dark blue lines, terminal exons. Colored dots, introns: blue, phase 0, inserted between two codons; green, phase 1, inserted between the first and second codon nucleotide; red, phase 2, inserted between the second and third codon nucleotide. Mating type allele shown, for those species with multiple sequenced mating types: *T. thermophila*: mt II; *T. borealis*: mt X; *T. pigmentosa*: mt III; *T. americanis*: mt X. Green triangle, phase 1 intron that exists in

Figure 3. *Continued*
T. borealis MTAY (but not in *MTAX* and *MTAZ*). Note that mating type genes in the asexual species, *T. pyriformis*, *T. vorax* and *T. paravorax*, are essentially silent during growth and starvation (see Figure 2C) and therefore lack a reliable RNA-Seq-based intron/exon annotation for most exons.

Interestingly, conserved intron #4 is absent from the *MTA* genes of all sequenced mating types in the "The-Mal" subgroup. These genes all have another intron (intron #6.5), located between conserved introns #6 and #7. Unexpectedly, the *T. borealis MTAY* gene contains an intron at exactly this location (#6.5) and phase (Figure 3B, green triangle and Table S1) but also has intron #4. Conceivably, the *MTA* genes of the "The-Mal" subgroup and the *T. borealis MTAY* gene may share a recent common ancestor.

Paramecium, like *Tetrahymena*, belongs to the Ciliate class Oligohymenophorea. The two genera are estimated to have diverged from one another nearly a billion years ago (Xiong et al., 2019). The mating protein (*mtA*) in *Paramecium tetraurelia* is also a member of the superfamily of genes having a terminal exon containing "Furin-like repeat" domains, which conserves all the cysteines found in *Tetrahymena* and the transmembrane helices (Singh et al., 2014). However, the *Paramecium* gene has only three introns, all at different locations than in the *Tetrahymena* genes. Thus, the *Paramecium* and *Tetrahymena* mating type genes likely had a common ancestor but have undergone extensive independent evolution.

The *Tetrahymena* mating type proteins exhibit a special type of incomplete lineage sorting

The topology of the individual MTA and MTB branches of the phylogenetic tree of the two entire proteins (Figure S5) does not exactly match the topology of the species tree (Figure 2B). This finding represents an example of "incomplete lineage sorting". The discrepancy is limited to the MTA and MTB proteins of *T. malaccensis*, *borealis*, *canadensis*, and *shanghaiensis*.

Interestingly, the topologies of the branches of both the MTA and MTB proteins in the phylogenetic tree for the distal third (C-terminal exon) (Figure 4A) are almost identical to the topology of the species tree (Figure 2B). Indeed, essentially all of the incomplete lineage sorting seen for the whole proteins is accounted for by that in the proximal (N-terminal) roughly two thirds of each protein (Figure 4B). In this context, it is important to note the very different functions of the two segments of the mating type proteins. The distal third (encoded by the C-terminal exon) includes the only predicted intracellular segment of the protein and can thus be inferred to be involved in mating type-non-specific interactions with cell machinery required, for example, for the structural remodeling of the cell in preparation for mating: re-shaping and de-ciliation of the anterior ventral surface where two cells will form a temporary junction lasting many hours ("tip transformation" (Wolfe and Grimes, 1979)). This is the protein segment that has evolved at a rate commensurate with that of other conserved, mating type-unrelated cell proteins. On the other hand, the proximal two thirds of the mating type proteins are predicted to be extracellular and to be the main site of the mating type-specific (positive and negative) interactions that allow the self vs. non-self recognition required to initiate or inhibit mating between two cells. Given that most *Tetrahymena* species possess multiple mating type systems, and that speciation has been accompanied by conservation of some mt protein specificities and evolutionary radiation of others (as described in subsequent sections), the incomplete lineage sorting observed for this protein segment becomes readily understandable. This clear distinction between nearly complete lineage sorting in the distal third and significantly incomplete lineage sorting in the proximal two thirds, seen for both of the two proteins, represents an example of what could be called "composite lineage sorting".

The individual examples of incomplete lineage sorting detected in this proximal two thirds of the MTA and MTB proteins are addressed in more detail below:

1) *T. malaccensis* MTAX and MTBX proteins co-branch with *T. thermophila* MTA4 and MTB4, respectively. The two species are very closely related (Figure 2B). This was already reported (Cervantes et al., 2013) and interpreted to mean that these two mtGPs recently evolved from the same mtGP in a common ancestor of the "The-Mal" subgroup.
2) *T. borealis* MTAZ and MTBZ proteins co-branch with *T. canadensis* MTAX and MTBX, respectively, which has a proposed analogous explanation to the previous case. These two species are among the most closely related *Tetrahymena* species pairs known. Their SSUrRNA genes are identical, and their *COX1* barcodes show only 4.4% polymorphisms (data not shown). Four percent is the

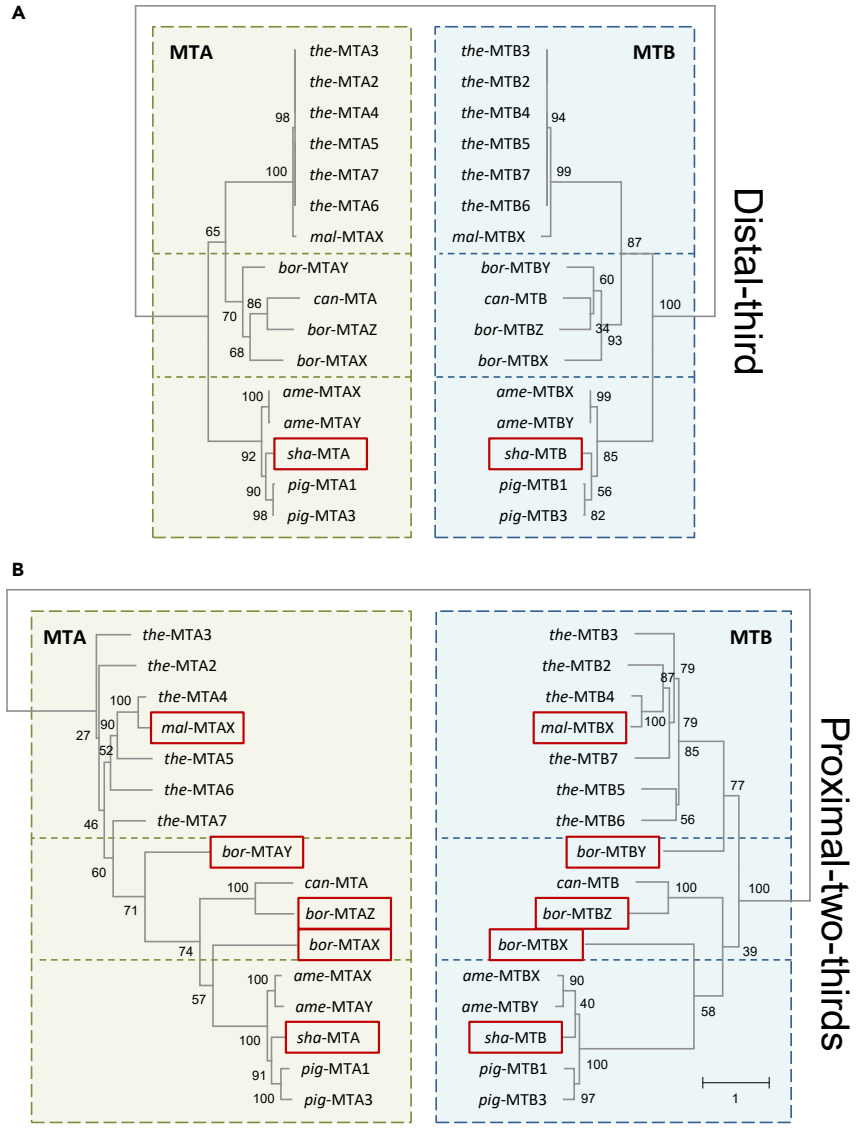


Figure 4. Phylogenetic tree of mating type proteins

(A) Protein phylogenetic tree based on the C-terminal exon (distal-third).

(B) Phylogenetic tree based on the rest of sequence (proximal-two-thirds). The best-fit models were calculated by ProtTest (version 3.4.2) (Darriba et al., 2011). (A) is under JTT + I + G + F model; (B) is under VT + I + G + F model. Numbers at each node, bootstrap values (1000 replicas). Branch length, number of base substitutions per site. Horizontal dashed lines: boundaries between the species clades/subgroups. Red boxes: deviations from the species tree. The MTA and MTB branches are shown opposite to one another to facilitate comparisons between them.

COX1 threshold that best corresponds to the ultimate criterion of the *Tetrahymena* species difference, the failure to mate (Doerder, 2019).

The two other discrepancies are more intriguing.

3) The co-branching of *T. borealis* mating type Y with all the *T. thermophila* and *T. malaccensis* mating types is unexpected because these species are in two different phylogenetic subgroups ("Bor-Can" and "The-Mal", respectively). This finding suggests that this particular mtGP has been retained and has changed relatively little since the divergence of the two subgroups. Further supporting this hypothesis, and as described in an earlier section, the *T. borealis MTAY* gene contains an intron at

exactly the same location and phase as the additional intron (#6.5) found in all the sequenced *MTA* genes in the "The-Mal" subgroup and nowhere else (Figure 3B).

4) The last anomaly is the co-branching of the MTA and MTB proteins of the perpetual selfer *T. shanghaiensis* with *T. pigmentosa* MTA3 and MTB3, respectively; this finding is addressed in the *T. shanghaiensis* section, further below.

A better understanding of the molecular basis of these cases of incomplete lineage sorting will require sequencing additional mating type genes of these and other species in the "Bor-Can" subgroup, as well as additional knowledge of the sequence and organization of the genes encoding these proteins in the germline (micronuclear) *mat* locus. The latter information is currently available only for *T. thermophila* (Cervantes et al., 2013).

MTA and *MTB* genes have coevolved within the different *Tetrahymena* phylogenetic subgroups

T. thermophila MTA and *MTB* gene products have non-redundant functions required for mating (Cervantes et al., 2013). Intriguingly, when we compared the two mating type genes in species belonging to different *Tetrahymena* phylogenetic groups, we noticed several cases where recent evolutionary changes in the *MTA* gene have been mirrored by corresponding changes in the *MTB* gene. For example, in the "The-Mal" subgroup, the GC content of 3'-terminal exons of both genes is significantly higher than that of other regions of the mtGP (Figures 5A and 5B). This difference is not observed in the other subgroups.

More striking evidence of coevolution is the length of mating type specific region. Previous work in *T. thermophila* (Cervantes et al., 2013) had revealed that the 3'-terminal exons in *MTA* genes, comprising about 1/3 of each gene, are highly conserved among alleles for the different mating types, while the rest of each gene is mating type-specific; the same is true for *MTB* alleles (Figure 5C, *T. the*). We examined different mtGPs in species of the "Bor-Can" subgroup (*T. borealis*) and the "Australis" clade (*T. americanis* and *T. pigmentosa*) to see if they also shared a distinct sequence conservation boundary. Sequence conservation plots (Figure 5C) of the *MTA* and *MTB* genes of these species show that none of their mating type genes have an abrupt conservation boundary at conserved intron #7. Instead, in the two species of the "Australis" clade, the conserved regions of the *MTA* and *MTB* genes are about twice as long as in *T. thermophila* (Figure 5C, *T. ame* and *T. pig*), so that only the 5'-terminal ~1/3 of each gene is unique for each mating type. In further contrast, for the three sequenced mating types of *T. borealis*, essentially the entire length of the *MTA* and *MTB* genes is mating type-specific (Figure 5C, *T. bor*).

In an attempt to shed more light on the question of MTA and MTB protein coevolution, we also did an amino acid usage analysis of all the MTA and MTB proteins in the sexual species (Data S1, Figure S6). The results showed some regularities but did not provide clear conclusions. A rigorous answer to this question will likely have to wait for additional experimental investigations and knowledge of the 3D structure of these proteins.

T. shanghaiensis cells have one mtGP allele and mate with one another

In contrast to the multiple mating type systems of other *Tetrahymena* species, *T. shanghaiensis* has been reported to be a selfer species by Chen et al. (1982) and Feng et al., 1988, in which sexually mature cells within every *T. shanghaiensis* clone mate with one another upon starvation. By DNA sequencing, we identified a single mtGP in the *T. shanghaiensis* MAC genome, whose *MTA* and *MTB* genes are homologous with the respective mating type genes of the other *Tetrahymena* species investigated here (Figures S2 and S4). Furthermore, the *T. shanghaiensis MTA* and *MTB* homologs have identical expression profiles to those of all the other sexual *Tetrahymena* species during growth and starvation Figure 2C.

To verify the previously reported observation that the non-assorting selfing trait is transmitted to sexual progeny, we did RNA-Seq experiments and *de novo* sequence assembly on two additional starved *T. shanghaiensis* populations, obtained as sexual progeny of independent selfing populations (i.e. biological replicates) (Figure S7, frames 5 and 7, highlighted with red stars). Cells in both populations contained transcripts from the *MTA* and *MTB* genes of the only previously detected single, genomic mtGP.

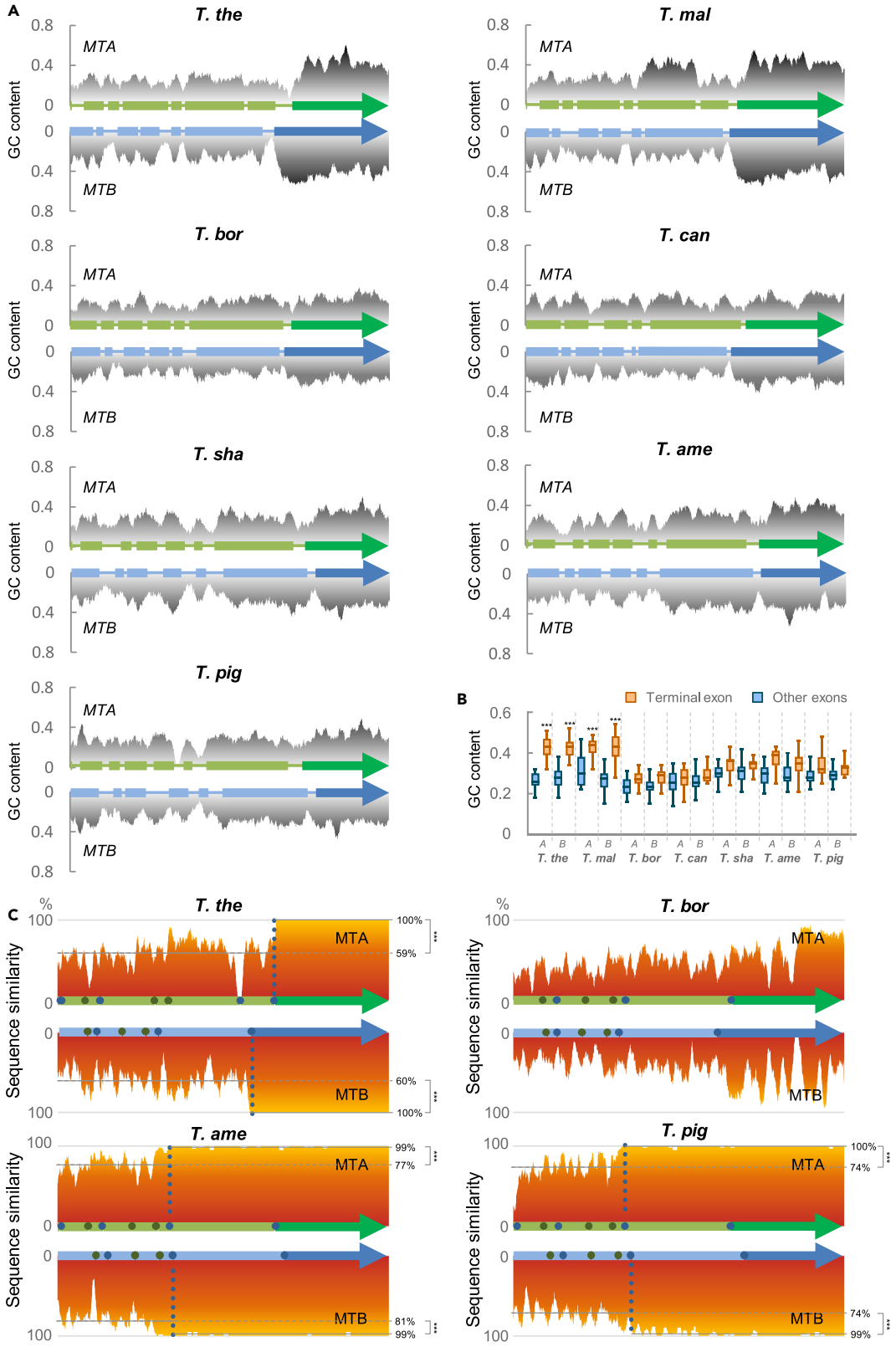

A
T. the
T. mal
T. bor
T. can
T. sha
T. ame
T. pig
MTA
MTB
GC content
0.8
0.4
0
B
Terminal exon
Other exons
GC content
0.6
0.2
A B
C
Sequence similarity
%
100
100%
59%
60%
99%
77%
81%
74%

Figure 5. Coevolution of the *MTA* and *MTB* genes as demonstrated by GC content and length of conserved regions

MTA (green) and *MTB* (blue) genes of the mtGP are shown parallel to each other to facilitate comparison.

(A) mtGPs structure and GC content. Dark green and dark blue exons, terminal exons of *MTA* and *MTB* genes, respectively.

(B) Statistical comparison (ANOVA) of the data shown in panel (A) Only exon sequence was included to avoid intron sequence influence. This analysis confirms that the GC content of terminal exons relative to that of other exons is statistically significantly higher in *T. thermophila* and *T. malaccensis*, and only in those to species.

(C) Sequence conservation among different mating type genes within species with more than one sequenced mtGP. Dots within the gene lines: introns (blue, phase 0; green, phase 1). Vertical dotted line: boundary between conserved and specific regions. Horizontal dashed lines: mean allele sequence similarity within each region. Specific regions and statistical comparisons (t-test) are shown on the right in each chart. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$. Note the co-variation in the length of highly conserved sequence in the *MTA* and *MTB* genes of the same species.

To explain the selfing of *T. shanghaiensis*, we proposed four *a priori* hypotheses, the first three of which differ with respect to what is encoded in the MAC genome, as illustrated in Figure S8.

1) Two normal mtGPs with different mating type specificities are present in the MAC of every cell; these mtGPs cannot be purified by assortment because the MAC mat locus is homozygous for both genes. MTA and MTB proteins with different mating type specificities are expressed upon starvation and trigger selfing.
2) Two normal mtGPs with different mating type specificities are present in the homozygous MAC genome but only one can be expressed. Frequent gene conversion causes a random mtGP to be expressed in every cell. Thus, even a clonal population will have a mixture of cells expressing different mating type specificities, leading to selfing upon starvation.
3) Only one type of mtGP is present in the homozygous MAC genome, containing *MTA* and *MTB* genes of different mating type specificity, i.e. a heterotypic mtGP. The proteins encoded by this mtGP are sufficient to trigger mating upon starvation.

Our finding of a single mtGP in every population is consistent with only the third hypothesis, a heterotypic mtGP.

It could be argued that mating in *T. shanghaiensis* is controlled by genes other than the mtGP found in the genome. This alternative is unlikely for the following reasons. As already mentioned, the *T. shanghaiensis MTA* and *MTB* genes show identical expression pattern during the life cycle as *MTA* and *MTB* genes of other *Tetrahymena* species (Figure 2C). Furthermore, the locations of both genes in the mating type protein phylogenetic tree (Figure 4) and in the amino acid frequency-based clustering (Figure S6) all correspond closely with the location of *T. shanghaiensis* in the species phylogeny tree (Figure 2B). Therefore, it seems highly probable that the mtGP of *T. shanghaiensis* functions in mating. Thus the results are entirely compatible with *T. shanghaiensis* having a heterotypic mtGP which is responsible for the perpetual selfing, as shown in Figure S8C.

The above conclusion is supported by the findings of Lin and Yao (2020) in *Tetrahymena thermophila*, published while this article was under review. These authors report that starved cells expressing one complete *MTA* gene of one mt specificity and one complete *MTB* gene of different mt specificity behave as non-assorting selfers, exactly as *T. shanghaiensis*, regardless of which pair of different mt specificities are involved. This finding strengthens our conclusion that the basis for the non-assorting selfing of *T. shanghaiensis* is the possession of *MTA* and *MTB* genes of different specificity. The non-assorting *T. shanghaiensis* selfer and the selfers investigated by Lin & Yao illustrate a new type of molecular basis for selfing in *Tetrahymena* (see Data S2 for more details about classification of *Tetrahymena* selfers). Rigorous proof of our hypothesis must await the development of molecular genetic tools to experimentally modify mating type genes in *T. shanghaiensis*.

Discussion

Proposed steps in the evolution of the *Tetrahymena* mtGP

Two mating type proteins (MTA and MTB) embody mating type specificity in the multiple mating type system of *T. thermophila*. These proteins contain ~1,500 amino acids each, and are encoded by adjacent head-to-head genes in the mtGP (Cervantes et al., 2013). The two genes have very similar organization and both belong to the superfamily of FLR domain proteins, but they only share about 6% sequence similarity. In this report, we identified and compared the MAC mating type protein-coding loci in nine

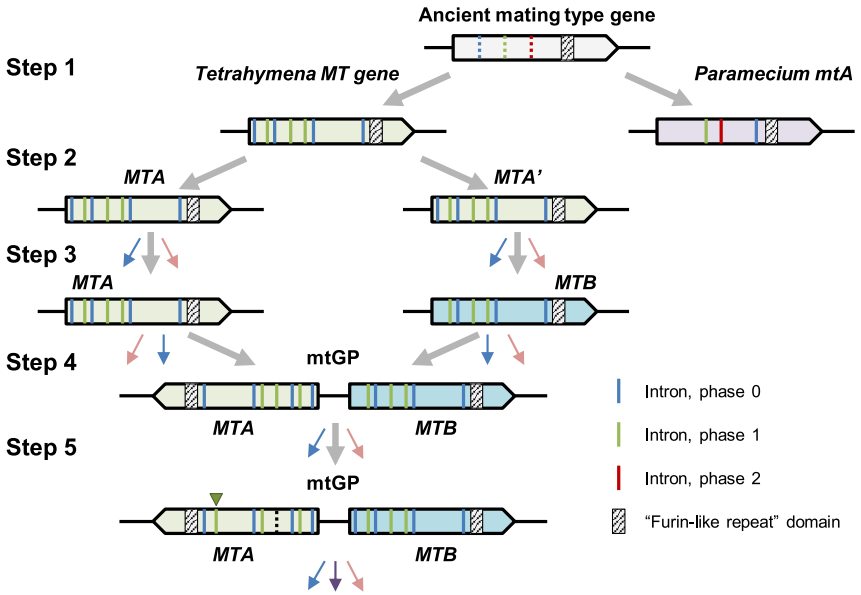


Figure 6. Postulated steps in the evolution of the *Tetrahymena* mtGP

Key to symbols: Thick gray arrows: major evolutionary steps. Color-matched thin arrows: paralogs encoding different mating type specificities (not shown) co-evolve from *MTA* and *MTB* genes. Symbols within each gene: vertical lines: introns (blue, green and red: phase 0, 1 and 2, respectively); hatched rectangle, "Furin-like repeat" domain. Step 1. Ancient gene, containing a cysteine-rich "Furin-like repeat" domain and five transmembrane helixes, acquires mating-related function, and evolves into distinct mating type genes in *Tetrahymena* and *Paramecium*. Step 2. In *Tetrahymena*, a DNA rearrangement generates a copy of the mating type gene (*MTA'*). Step 3. The *MTA'* gene acquires a new mating-related function, becoming the *MTB* gene. A set of paralogs encoding different mating type specificities (colored arrows) co-evolve from each gene. Intron location and phase are conserved during this differentiation and following events. Step 4. A second translocation causes *MTA* and *MTB* to become adjacent genes in opposite orientation, thus generating the head-to-head mtGP. A new wave of co-evolutionary mutational changes generates additional mating type specificities. Step 5. An *MTA* rearrangement deletes intron #4 (black vertical dotted line) and creates intron #6.5 (marked with triangle) within an *MTA* gene in a "The-Mal" subgroup ancestor, followed by a new wave of co-evolved changes that generate additional mating type specificities. Note: if the *MTA* and *MTB* alleles brought into contiguity at step 4 had encoded different mating type specificities, this evolutionary intermediate initially would have been a perpetual selfer, like present day *T. shanghaiensis*.

additional *Tetrahymena* species, encompassing the phylogenetic diversity of the genus. Eight of the additional species were found to have homologs to *T. thermophila MTA* and *MTB*, while an asexual strain of the most distantly related species, *T. paravorax*, has an *MTA* homolog but lacks any trace of an *MTB* homolog. The *MTA* and *MTB* genes of the sexual species share in common at least 5 introns that have conserved location and phase.

The findings we have reported here suggest the major steps in the evolution of the *Tetrahymena* mtGP over the last 150 Myr, illustrated in Figure 6:

1) A cysteine-rich "Furin-like repeat" domain protein acquired mating-related activity in an Oligohymenophorean ancestor of *Tetrahymena* and *Paramecium*. A terminal exon, encoding a "Furin-like repeat" domain and five transmembrane helixes, are shared by mating type genes of *P. tetraurelia* and all ten *Tetrahymena* species examined. The asexual *T. paravorax* strain, the earliest *Tetrahymena* to diverge, with a clearly differentiated *MTA*-like gene (*MTAL*) but lacking an adjacent *MTB* gene, could represent a relic of this evolutionary stage.
2) A duplication of the "Furin-like repeat" family protein gene occurred in a common ancestor of the "Australis" and "Borealis" *Tetrahymena* clades, and generated identical *MTA* and *MTB* genes sharing the location and phase of multiple introns still observed today.
3) Subsequent co-evolution of the duplicate copies generated functionally differentiated *MTA* and *MTB* genes.

4) A subsequent DNA rearrangement brought into contiguity two cognate *MTA* and *MTB* genes, in head-to-head orientation, generating the *Tetrahymena* mtGP that we see today.

5) A rearrangent in *MTA*, marked by the loss of intron #4 and the appearance of intron #6.5, occurred in an ancestor of the "The-Mal" subgroup.

It seems very likely that every step was quickly followed by waves of paralogous diversification of *MTA* and *MTB* to generate different mating type specificities.

It is possible that steps 2 and 4 occurred at once, i.e., step 2 was a "palindromic duplication", and thus the head-to-head *MTA-MTB* contiguity preceded their functional differentiation. However, this seems unlikely because intra-strand gene conversion within the palindrome would have precluded the functional differentiation of the two genes, due to mutual DNA sequence self-correction among the two copies, such as described in metazoan Y chromosome palindromic duplications (Trombetta and Cruciani, 2017).

These findings raise the question, what could have been gained by having *MTA* and *MTB* genes immediately adjacent to one another? The tight linkage of mating related genes provides important advantages, such as to "facilitate the coordinated expression" and "cosegregation of the interacting genes" (Uyenoyama, 2005). One additional consequence of *MTA-MTB* contiguity in *Tetrahymena* is that it minimizes the frequency of selfing among sexual progeny that would otherwise occur as a consequence of independent allelic assortment of the two genes in the MAC. If the *MTA* and *MTB* genes were located on different MAC chromosomes, the two genes would assort independently in double heterozygotes. MACs that are pure for non-cognate (heterotypic) *MTA* and *MTB* genes would then be frequently generated, ultimately resulting in non-assorting selfers (see Data S3 and Figure S9 for a detailed explanation). Thus, reducing the length and the sequence similarity of the intergenic segment between *MTA* and *MTB* genes has the effect of minimizing selfer-generating germline or somatic recombination events. The evolution of the mtGP, with its tight contiguity of the cognate (homotypic) *MTA-MTB* genes, likely was a major step in the evolution of the cross-breeding genetic economies generally observed in *Tetrahymena* species today.

Finally, the mtGP organization has also proven its versatility by allowing the evolution of additional genetic and molecular mechanisms of MTD capable of adjusting selfing frequency in *Tetrahymena*. These mechanisms can promote outbreeding or inbreeding under conditions of low or high reproductive stress, respectively (see below for details). A better understanding of these mechanisms will come when studies of the MIC organization of the mating type loci of the various species become available.

True-breeding selfing in *T. shanghaiensis*

Our work has confirmed that *T. shanghaiensis* is a true-breeding selfer and has shown that it contains a single mtGP in its MAC that behaves structurally and functionally like the mtGPs of the other species investigated here (Figures 2, 3, 4, S2, S4, and S6). The simplest explanation for its obligatory selfing behavior is that *T. shanghaiensis* mtGP is heterotypic, i.e. its *MTA* and *MTB* genes have different mating type specificity. This conclusion is supported by the finding that *T. thermophila* mutants, which fail to complete MTD and are left in the MAC with intermediates containing a complete *MTA* and a complete *MTB* gene but with different mt specificity, are non-assorting selfers (Lin and Yao, 2020) just like wild-type *T. shanghaiensis*. Evolutionarily, the *T. shanghaiensis* chimeric mtGP could have been generated by a simple DNA rearrangement, such as a non-homologous meiotic recombination event between two normal mtGPs of different mating types occurring at the *MTA-MTB* intergenic region in a heterozygote, resulting in the replacement of either gene with a homolog of different mating type specificity. An alternative way in which the heterotypic mtGP could have arisen is by successive mutations in the two genes of an initially homotypic mtGP that promoted increasingly strong interactions between their encoded proteins ultimately leading to efficient selfing, favored under conditions of high reproductive stress.

As a perpetual selfer, *T. shanghaiensis* can be considered to be unisexual, in the sense used to describe mating in the absence of any intra-species diversity at the mating type locus, as occurs in some fungi (reviewed by Heitman (2015)). Indeed, that author argues that the last eukaryotic common ancestor, was unisexual, i.e., a perpetual selfer using the *Tetrahymena* terminology. The species tree (Figure 2B) shows that *T. shanghaiensis* is surrounded by species with homotypic mtGPs. Thus, it seems most probable that the unisexuality (perpetual selfing) of *T. shanghaiensis* is derived, rather than ancestral in the genus

Tetrahymena. That does not exclude, however, the possibility of a unisexual common ancestor of the "Borealis" and "Australis" *Tetrahymena* clades.

The intron #4 and #6.5 rearrangement provide a glimpse into the evolution of a different mating type system

While sexual reorganization events (meiosis, fertilization, MAC differentiation) are highly conserved among Ciliates, a striking variety of mating type systems have evolved within this group (Phadke and Zufall, 2010). For example, in the multiple mating type system of the hypotrich Ciliate *Euplotes*, the two proteins that embody ligand and receptor function for each mating type specificity are the products of intron-splicing variants of the same gene; both proteins are very small, in the order of 40 amino acids, reminiscent of cytokines of multicellular eukaryotes (Luporini et al., 1986; Miceli et al., 1992). In the binary mating type system of the heterotrich Ciliate *Blepharisma*, the mating type ligand is not a protein but a small tryptophan-related molecule (Miyake, 1996; Sugiura et al., 2005). This diversity implies that the molecules that embody mating type specificity in Ciliates have independently undergone major successive replacements. However, the lack of evolutionary intermediates makes it extremely challenging to trace how this diversity evolved among major groups.

The results reported here have allowed us to infer a succession of replacement waves, occurring within the genus *Tetrahymena*, which generated a diversity of mating type proteins from an ancestral "Furin-like repeat" protein (Figure 6). Serendipitously, this work also uncovered a more recent replacement wave in which an *MTA* allele, generated along the way by the loss of conserved intron #4 and the gain of intron #6.5, is inferred to have *de novo* replaced the *MTA* alleles of mtGPs of every mating type specificity among the sequenced *MTA* genes in the "The-Mal" subgroup. Each of the rearrangements that resulted in the two intron changes likely happened only once, with the final rearrangement presumably resulting in one *MTA* gene of one particular mating type specificity. This variant *MTA* gene then had to spread and diversify, to ultimately be able encode every known *MTA* mating type specificity found today in the "The-Mal" subgroup. It seems reasonable to expect that *MTB* genes also had to co-evolve, in order to allow all the appropriate positive and negative mating type protein interactions required to promote mating between different mating types and prevent selfing in the multiple mating type system. The generation of "raw material" for the re-evolution of multiple mtGPs of different mating type—all containing the variant introns in the *MTA* gene—was likely facilitated by two special features: the tandem array organization of multiple mtGPs in the *T. thermophila* germline (micronuclear) genome (Cervantes et al., 2013), in combination with unequal meiotic crossing over, a capacity which has been well documented in the case of *Tetrahymena* leucine-rich repeat genes (Xiong et al., 2019).

The mtGP: a durable and effective vehicle for *Tetrahymena* unicell adaptation to reproductive stress fluctuations

T. thermophila cells, maintained by asexual reproduction in the laboratory for long periods in the absence of mating, eventually become sterile (Simon and Nanney, 1979). This failure was inferred to be due to the random accumulation of deleterious mutations in the MIC. The time-sensitive deterioration of their germline results in the susceptibility of *Tetrahymena* cells to reproductive stress.

Most *Tetrahymena* species tend to be primarily outbreeders. As previously discussed (Orias et al., 2017), this represents a balance between mechanisms that promote outbreeding and inbreeding. Features of *Tetrahymena* biology that promote outbreeding include a long sexual immaturity period, intranuclear coordination during MTD, allelic assortment and, in the "Australis" clade, synclonal MTD. Features capable of promoting inbreeding include multiple mating type systems, selfing, and karyonidal MTD in the "Borealis" clade. This investigation of the mating type genes of a broader set of *Tetrahymena* species has contributed an additional finding relevant to the balance between inbreeding and outbreeding, namely the rare occurrence of a putative heterotypic mtGP, which ensures obligatory, perpetual selfing in *T. shanghaiensis*.

Some asexual *Tetrahymenas*, such as *T. pyriformis* and *T. vorax* also studied here, have lost their MIC (so called amicronucleates, amics) and can no longer conjugate. Such an extreme feature avoids the germline deterioration that would affect sexual *Tetrahymena* cells under conditions of severe sexual reproductive stress. *Tetrahymena* amics are presumably capable of long-term adaptation to changing environments by virtue of allelic assortment in the polyploid MAC. The number of copies of favorable mutations can

increase by random assortment and come to phenotypic expression, while unfavorable mutations can be eliminated.

A puzzling feature of the asexual strains investigated here is that their mtGP retain open reading frames (at least for the sequenced exons), even though the proteins are no longer needed for mating, and their expression is not induced by starvation. One trait that would delay the emergence of internal in-frame stop codons is the variant Ciliate genetic code, which has a single stop codon, UGA. It is also possible that the mtGP has other useful function(s), unrelated to mating and expression is induced by some condition other than starvation, which keeps it under selection.

The species investigated here were chosen to provide a sample of the diversity of the *Tetrahymena* genus, which now contains nearly 100 identified species, with no end in sight. As the breeding systems of additional *Tetrahymena* species are molecularly characterized, our current picture of how they have evolved in reaction to fluctuating levels of reproductive stress will no doubt be enlarged and enriched.

Limitations of the study

In this work, we mainly focused on the MAC mtGP. Even though these results provide some clues of the evolution of the MIC mating type locus, there are still many unknowns. Subsequent studies of MIC mtGPs should provide a more elaborate picture of the evolutionary process. In addition, the most distantly related species, *T. paravorax*, seems to be asexual, so it will be more informative if we can find and investigate a sexual outgroup species in the future.

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Wei Miao (miaowei@ihb.ac.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The accession number for the data reported in this paper is GEO: PRJNA510545.

Methods

All methods can be found in the accompanying Transparent methods supplemental file.

Supplemental information

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101950.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant number 31525021 and 91631303 to W.M.), by the Wuhan Branch, Supercomputing Center, Chinese Academy of Sciences, China, and by grant of the Academic Senate of the University of California Santa Barbara (E.O.). The authors are extremely grateful to three anonymous reviewers, whose insightful questions and suggestions materially improved the quality of this article.

Author contributions

G.Y. designed and performed most of the experiments, analyses, and wrote the manuscript. W.Y. contributed to genome/transcriptome assembling. X.H. amplified and sequenced introns and intergenic regions of mtGPs which only have RNA-Seq data. K.C. contributed to mating type test and strain storage. J.X. and E.O. contributed to data analyses. E.H. and E.O. reviewed the manuscript. W.M. contributed to the experimental design, and reviewed the manuscript.

Declaration of interests

The authors declare no competing interests.

Received: February 25, 2020
Revised: June 16, 2020
Accepted: December 11, 2020
Published: January 22, 2021

References

Bachtrog, D., Mank, J.E., Peichel, C.L., Kirkpatrick, M., Otto, S.P., Ashman, T.L., Hahn, M.W., Kitano, J., Mayrose, I., Ming, R., et al. (2014). Sex determination: why so many ways of doing it? PLoS Biol. *12*, e1001899.

Cervantes, M.D., Hamilton, E.P., Xiong, J., Lawson, M.J., Yuan, D., Hadjithomas, M., Miao, W., and Orias, E. (2013). Selecting one of several mating types through gene segment joining and deletion in *Tetrahymena thermophila*. PLoS Biol. *11*, e1001518.

Chen, Y., Luo, Z., and Cao, T. (1982). Conjugation in *Tetrahymena pyriformis* S1, a selfer strain from Shanghai. Acta Zool. Sin. *28*, 319–324.

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics *27*, 1164–1165.

Doerder, F.P. (2014). Abandoning sex: multiple origins of asexuality in the ciliate *Tetrahymena*. BMC Evol. Biol. *14*, 112.

Doerder, F.P. (2019). Barcodes reveal 48 new species of *Tetrahymena, Dexiostoma*, and *Glaucoma*: phylogeny, ecology, and biogeography of new and established species. J. Eukaryot. Microbiol. *66*, 182–208.

Feng, S., Sun, Q., Cao, T., Li, L., and Chen, Y. (1988). The S1 strain of *Tetrahymena* from Shanghai – *Tetrahymena shanghaiensis.nov*. Acta Zool. Sin. *34*, 42–51.

Gruchy, D.F. (1955). The breeding system and distribution of *Tetrahymena pyriformis*. J. Eukaryot. Microbiol. *2*, 178–185.

Heitman, J. (2015). Evolution of sexual reproduction: a view from the Fungal Kingdom supports an evolutionary epoch with sex before sexes. Fungal Biol. Rev. *29*, 108–117.

Kües, U. (2015). From two to many: multiple mating types in Basidiomycetes. Fungal Biol. Rev. *29*, 126–166.

Lin, I.T., and Yao, M.C. (2020). Selfing mutants link Ku proteins to mating type determination in *Tetrahymena*. PLoS Biol. *18*, e3000756.

Luporini, P., Raffioni, S., Concetti, A., and Miceli, C. (1986). The ciliate *Euplotes raikovi* heterozygous at the mat genetic locus coreleases two individual species of mating pheromone: genetic and biochemical evidence. Proc. Natl. Acad. Sci. U S A *83*, 2889–2893.

Miceli, C., La Terza, A., Bradshaw, R.A., and Luporini, P. (1992). Identification and structural characterization of a cDNA clone encoding a membrane-bound form of the polypeptide pheromone Er-1 in the ciliate protozoan *Euplotes raikovi*. Proc. Natl. Acad. Sci. U S A *89*, 1988–1992.

Miyake, A. (1996). Fertilization and sexuality in ciliates. In Ciliates: Cells as Organisms, K. Hausmann and P.C. Bradbury, eds. (Gustav Fischer Verlag), pp. 1–25.

Orias, E. (1981). Probable somatic DNA rearrangements in mating type determination in *Tetrahymena thermophila*: a review and a model. Dev. Genet. *2*, 185–202.

Orias, E., Cervantes, M.D., and Hamilton, E.P. (2011). *Tetrahymena thermophila*, a unicellular eukaryote with separate germline and somatic genomes. Res. Microbiol. *162*, 578–586.

Orias, E., Singh, D.P., and Meyer, E. (2017). Genetics and epigenetics of mating type determination in *Paramecium* and *Tetrahymena*. Ann. Rev. Microbiol. *71*, 133–156.

Phadke, S.S., and Zufall, R.A. (2010). Rapid diversification of mating systems in ciliates. Biol. J. Linn. Soc. *98*, 187–197.

Sela, M., and Lifson, S. (1959). On the reformation of disulfide bridges in proteins. Biochim. Biophys. Acta *36*, 471–478.

Simon, E.M. (1980). Mating-type inheritance and maturity times in crosses between subspecies of *Tetrahymena pigmentosa*. Genetics *94*, 93–113.

Simon, E.M., and Nanney, D.L. (1979). Germinal aging in *Tetrahymena thermophila*. Mech. Ageing Dev. *11*, 253–268.

Simon, E.M., Nanney, D.L., and Doerder, F.P. (2009). The "*Tetrahymena pyriformis*" complex of cryptic species. In Protist Diversity and Geographical Distribution, W. Foissner and D.L. Hawksworth, eds. (Springer Netherlands), pp. 131–146.

Singh, D.P., Saudemont, B., Guglielmi, G., Arnaiz, O., Gout, J.F., Prajer, M., Potekhin, A., Przybos, E., Aubusson-Fleury, A., Bhullar, S., et al. (2014). Genome-defence small RNAs exapted for epigenetic mating-type inheritance. Nature *509*, 447–452.

Sugiura, M., Kawahara, S., Iio, H., and Harumoto, T. (2005). Developmentally and environmentally regulated expression of gamone 1: the trigger molecule for sexual reproduction in *Blepharisma japonicum*. J. Cell Sci. *118*, 2735–2741.

Thornton, J.M. (1981). Disulphide bridges in globular proteins. J. Mol. Biol. *151*, 261–287.

Trombetta, B., and Cruciani, F. (2017). Y chromosome palindromes and gene conversion. Hum. Genet. *136*, 605–619.

Uyenoyama, M.K. (2005). Evolution under tight linkage to mating type. New Phytol. *165*, 63–70.

Wolfe, J., and Grimes, G.W. (1979). Tip transformation in *Tetrahymena*: a Morphogenetic response to interactions between mating types. J. Protozool. *26*, 82–89.

Xiong, J., Yang, W., Chen, K., Jiang, C., Ma, Y., Chai, X., Yan, G., Wang, G., Yuan, D., Liu, Y., et al. (2019). Hidden genomic evolution in a morphospecies—the landscape of rapidly evolving genes in *Tetrahymena*. PLoS Biol. *17*, e3000294.

iScience, Volume *24*

Supplemental Information

Evolution of the mating type gene pair and multiple sexes in *Tetrahymena*

Guanxiong Yan, Wentao Yang, Xiaojie Han, Kai Chen, Jie Xiong, Eileen P. Hamilton, Eduardo Orias, and Wei Miao

Supplemental Information

Figure S1. Mating type determination in *Tetrahymena*. Related to Figure 1. (**A**) Synclonal MTD. The four karyonides from a pair are usually determined by Mendelian genetics to express the same mating type (shown as the same color) which could restrict inbreeding. All possible outcomes (genotypes and mating types) are illustrated in the sample cross shown. (**B**) Karyonidal MTD. The four karyonides from a pair express independently determined mating types, unrelated to the mating type of either parent or to those of the other karyonides from same pair. Thus, karyonidal MTD could favor inbreeding. For more information, please see Phadke and Zufall (2010) and Orias et al. (2017).

Figure S2. Sequence alignment and conservation of intron location and phase among all sequenced MTA, MTB and MTAL *Tetrahymena* proteins. Related to Figure 3B. Colored dots, introns: blue, phase 0, inserted between two codons; green, phase 1, inserted between the 1st and 2nd codon nucleotide; red, phase 2, inserted between the 2nd and 3rd codon nucleotide). Black lines separate MTA, MTB and MTAL proteins. “Dark red background”, identical site (Global score = 1); “Pink letter on white background”, high similarity (Global score over 0.7); “Black letter on white background”, low similarity (Global score below 0.7). Similarity at each site was based on GONNET matrices. “Pink background”, conserved cysteine residue in all proteins except MTAL.

Figure S3. mtGP synteny in *Tetrahymena* species. Related to Figure 2A. Note that a) the *MTAL* gene of *T. paravorax* is syntenic with all the other *MTA* genes; but there is no *MTB* gene homolog; b) the synteny of *T. americanis* mtGP could not be tested due to lack of a sequenced genome and c) *T. vorax* is the only species whose mtGP not syntenic with others.

Figure S4. Strong conservation of *MTA* and *MTB* intron location and phase within each of the three subgroups. Related to Figure 3B. Intron colors as in Fig. S2. Only relevant segments of the multiple alignments are shown in this figure.

Figure S5. Phylogenetic tree of mating type proteins based on full length MTA and MTB sequences. Related to Figure 4. The phylogenetic tree was constructed under WAG+I+G+F model. Numbers at each node, bootstrap values (1000 replicas). Branch length, number of base substitutions per site. Red boxes: deviations from the species tree. The MTA and MTBs branches are shown opposite to one another to facilitate comparisons between them.

Figure S6. MTA and MTB proteins show different amino acid usage frequencies among different subgroups. Related to Figure 5. (**A**), (**B**) and (**C**) Protein clustering based on Euclidean distance calculated by amino acid usage frequency. (**A**) is based on full-length sequences; (**B**) is based on C-terminal exon (Distal-third) encoded sequences; (**C**) is based on the rest of sequence (Proximal-two-thirds). Color bar, amino acid usage frequency. Min and Max, minimum and maximum value of each column, respectively. (**D**) Principal Components Analysis (PCA) of amino acid usage frequency (based on full-length sequences). Data points for mating type proteins of each subgroup are shown in the same colors as in panel A. *T. shanghaiensis* data points are highlighted for reasons explained in the last Results section.

Figure S7. Pedigree of two *T. shanghaiensis* populations whose mtGPs were independently sequenced as biological replicates. Related to Figure 6. Key to panel labels. S, starvation. G, growth. Note that selfing leads to the same conjugation events described in **Figure 1**, including the generation of sexual progeny with new MICs and MACs. Green dashed oval in panel 4: mature progeny cell, originally derived from the pair within the green dashed circle in panel 3, which was cloned and allowed to starve and self again (panel 7). Red stars: two selfing cultures which represent independently sequenced biological replicates.

Figure S8. Possible mechanisms responsible for obligatory selfing in *T. shanghaiensis*. Related to Figure 6. Key to symbols: orange and blue, two different mating type specificities; divergent thick arrows, MAC mtGP; sticks attached to cell membrane, mating type proteins: thick, MTA; thin, MTB. (**A**) mtGPs with different mating type specificities are present in the MAC of every cell; these mtGPs cannot be purified by assortment because the MAC *mat* locus is homozygous for both alleles. MTA and MTB proteins with different mating type specificities are expressed upon starvation and trigger selfing. (**B**) mtGPs with two different mating type specificities are present in the homozygous MAC genome but only one can be expressed. Frequent gene conversion causes a random mtGP to be expressed in every cell. Thus, even a clonal population will have a mixture of cells expressing different mating type specificities, leading to selfing upon starvation. (**C**) Only one type of mtGP is present in the homozygous MAC genome, containing *MTA* and *MTB* genes of different mating type specificity, i.e. heterotypic mtGP. The proteins encoded by this mtGP are sufficient to trigger mating upon starvation.

Figure S9. *MTA* and *MTB* genes could have been on the same or different MAC chromosomes during mtGP evolution. Related to Figure 6. (**A**). The *MTA* and *MTB* genes are adjacent on the same MAC chromosome, the arrangement we find today. (**B**) and (**C**)**.** Possible types of *MTA* and *MTB* gene organization at stage 2 of mtGP evolution in **Figure 6**. The *MTA* and *MTB* genes are on <u>different</u> (**B**) or <u>same</u> (**C**) MAC chromosomes. **Key to symbols.** Left-pointing triangles, *MTA* genes; right-pointing triangles, *MTB* genes; blue and red, different mating type specificities (e.g., mt I and II). **Key to stages.** MAC initially heterozygous for I and II mating type specificities: **1**, before assortment; **2**, after assortment is completed; **3**, behavior of sexually mature terminal assortants after starvation (selfing or no selfing). **Under case A**, only a single, homotypic *MTA* and *MTB* gene pair remains in the MAC of terminal assortants (stage **2**) and no selfing is expected. **Under case B**, only a single *MTA* allele and a single *MTB* allele remain in the MAC of terminal assortants. The MACs of 50% of the terminal assortants will be pure for heterotypic *MTA* and *MTB* alleles, and those cells will be non-assorting selfers (see **Text S2**). **Under case C**, two main types of terminal assortants are expected with equal frequencies having homotypic *MTA* and *MTB* gene combinations, respectively. Rare MAC recombination could occasionally generate a heterotypic mtGP (bottom chromosome in panel **C1**); terminal assortants pure for this chromosome would be non-assorting selfers. (See **Text S3** for more details).

Table S1. Conservation of intron phase and location among functional *Tetrahymena* mating type genes. Related to Figure 3.

Table S2. Number of independent mating tests done within and between sexual *Tetrahymena* strains used. Related to Table 2.

Table S3. Collection sites and other information on previously unreported strains. Related to Table 2.

Data S1. Evolution of amino acid usage among the mating type genes of sexual *Tetrahymena* species. Related to Figure 5.

Data S2. Classification of *Tetrahymena* selfers. Related to Figure 6.

Data S3. Evolution of *MTA-MTB* contiguity after functional differentiation would have discouraged selfing. Related to Figure 6.

Figure S1

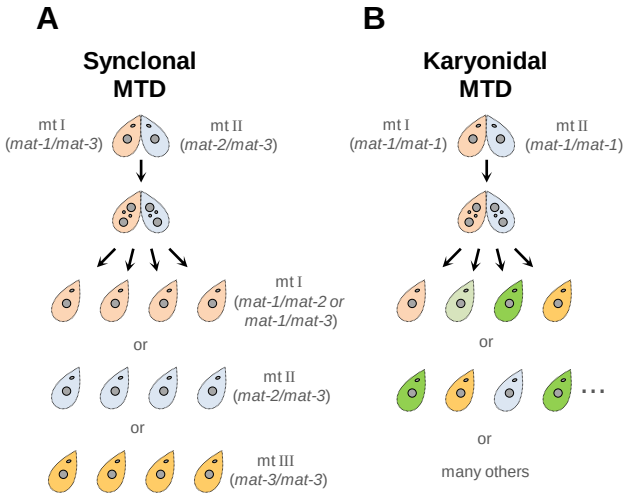

A
Synclonal
MTD
mt I
(mat-1/mat-3)
mt II
(mat-2/mat-3)
mt I
(mat-1/mat-2 or
mat-1/mat-3)
or
mt II
(mat-2/mat-3)
or
mt III
(mat-3/mat-3)
B
Karyonidal
MTD
mt I
(mat-1/mat-1)
mt II
(mat-1/mat-1)
or
or
many others

Figure S2

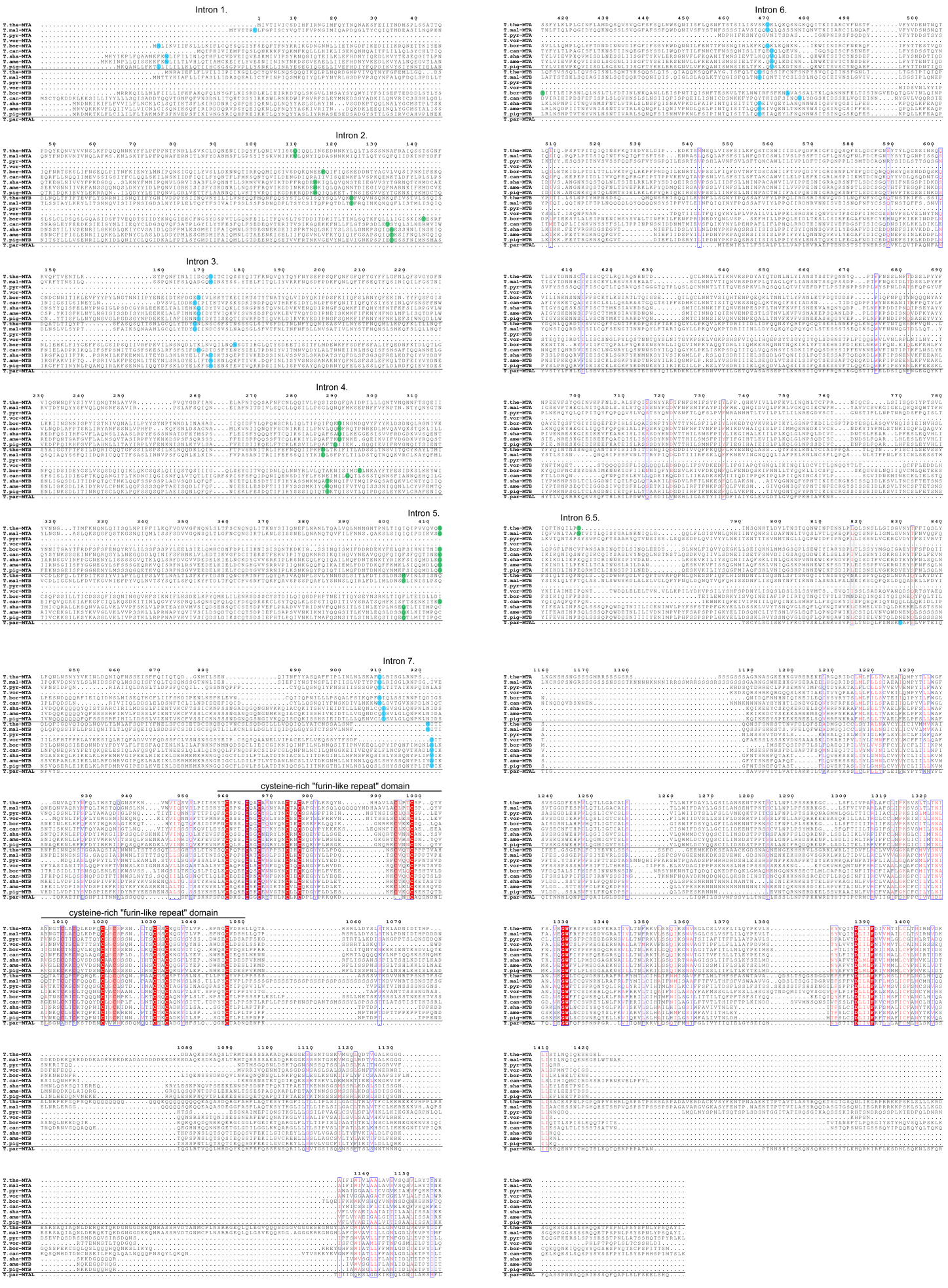

Intron 1.
Intron 6.
Intron 2.
Intron 3.
Intron 4.
Intron 5.
Intron 6.5.
Intron 7.
cysteine-rich "furin-like repeat" domain
cysteine-rich "furin-like repeat" domain

Figure S3

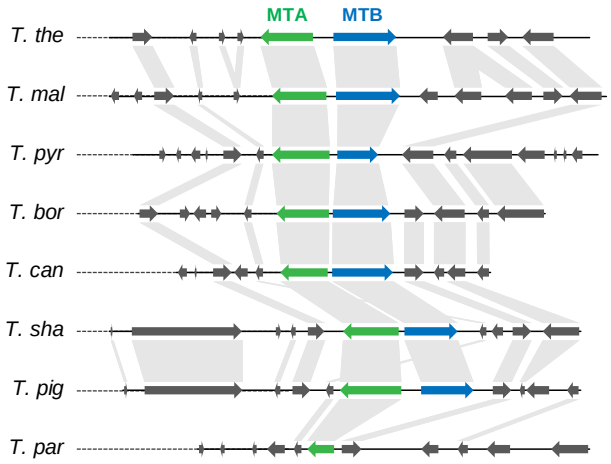


Figure S4

Figure S5

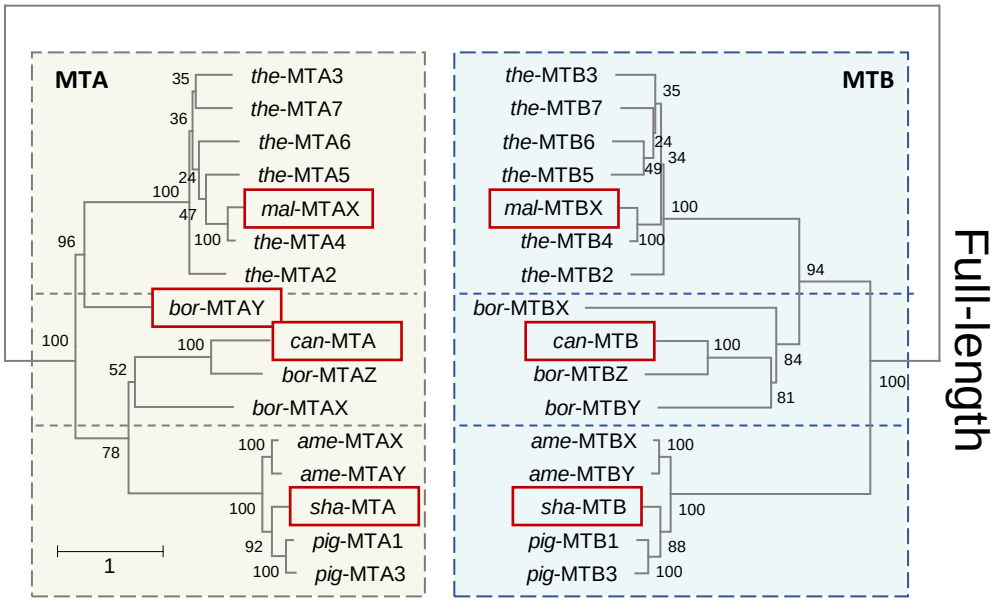

MTA
MTB
Full-length
the-MTA3
the-MTA7
the-MTA6
the-MTA5
mal-MTAX
the-MTA4
the-MTA2
bor-MTAY
can-MTA
bor-MTAZ
bor-MTAX
ame-MTAX
ame-MTAY
sha-MTA
pig-MTA1
pig-MTA3
1
the-MTB3
the-MTB7
the-MTB6
the-MTB5
mal-MTBX
the-MTB4
the-MTB2
bor-MTBX
can-MTB
bor-MTBZ
bor-MTBY
ame-MTBX
ame-MTBY
sha-MTB
pig-MTB1
pig-MTB3

Figure S6

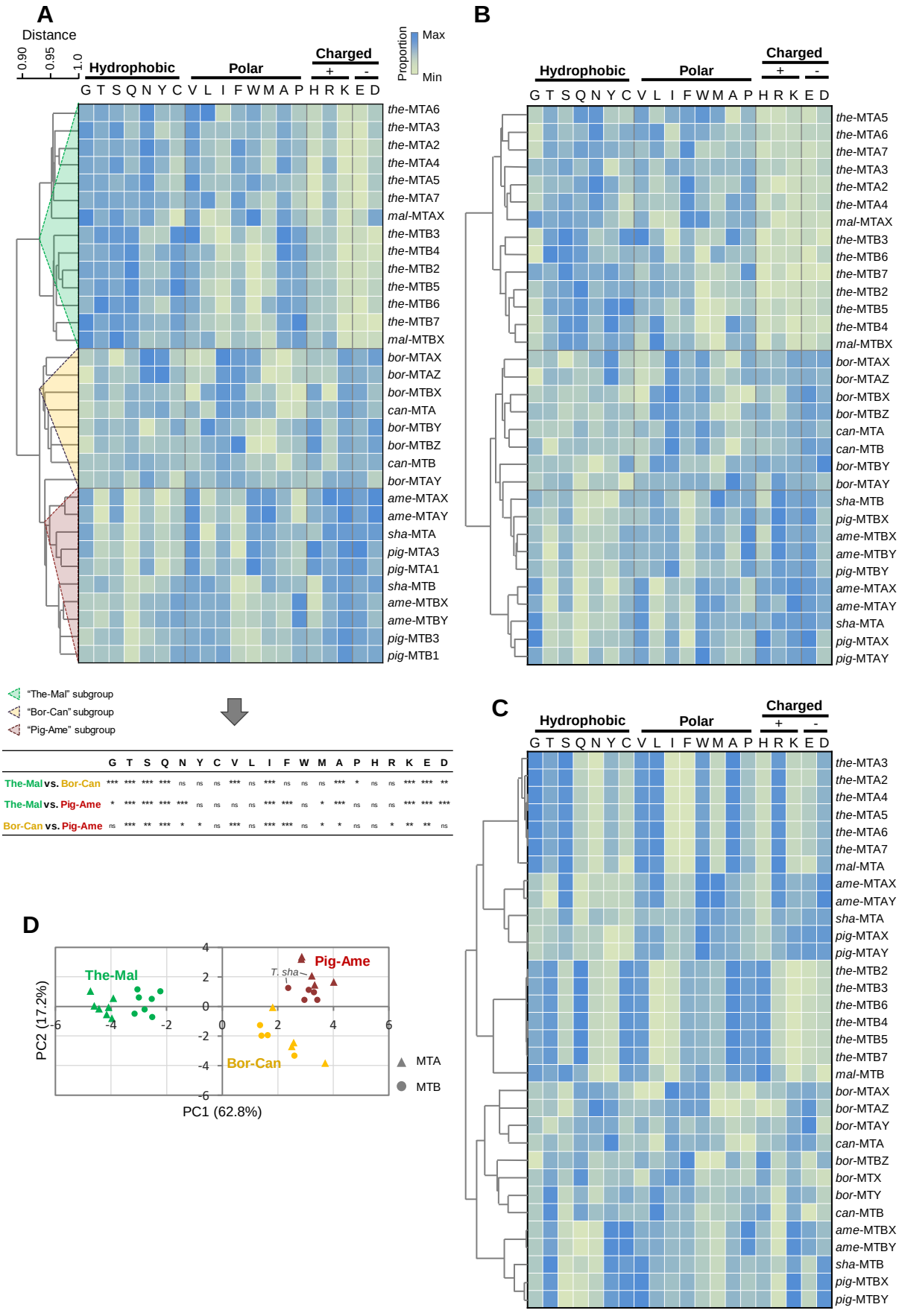

A
Distance
0.90
0.95
1.0
Hydrophobic
Polar
Charged
+
-
Proportion
Max
Min
G T S Q N Y C V L I F W M A P H R K E D
the-MTA6
the-MTA3
the-MTA2
the-MTA4
the-MTA5
the-MTA7
mal-MTAX
the-MTB3
the-MTB4
the-MTB2
the-MTB5
the-MTB6
the-MTB7
mal-MTBX
bor-MTAX
bor-MTAZ
bor-MTBX
can-MTA
bor-MTBY
bor-MTBZ
can-MTB
bor-MTAY
ame-MTAX
ame-MTAY
sha-MTA
pig-MTA3
pig-MTA1
sha-MTB
ame-MTBX
ame-MTBY
pig-MTB3
pig-MTB1
"The-Mal" subgroup
"Bor-Can" subgroup
"Pig-Ame" subgroup
G T S Q N Y C V L I F W M A P H R K E D
The-Mal vs. Bor-Can *** *** *** *** ns ns ns *** ns *** ns ns *** * ns ns *** *** **
The-Mal vs. Pig-Ame * *** *** *** *** ns ns ns ns *** *** ns * *** ns ns ns *** *** ***
Bor-Can vs. Pig-Ame ns *** ** *** * * ns *** ns *** *** ns * * ns ns * ** ** ns
B
Hydrophobic
Polar
Charged
+
-
G T S Q N Y C V L I F W M A P H R K E D
the-MTA5
the-MTA6
the-MTA7
the-MTA3
the-MTA2
the-MTA4
mal-MTAX
the-MTB3
the-MTB6
the-MTB7
the-MTB2
the-MTB5
the-MTB4
mal-MTBX
bor-MTAX
bor-MTAZ
bor-MTBX
bor-MTBZ
can-MTA
can-MTB
bor-MTBY
bor-MTAY
sha-MTB
pig-MTBX
ame-MTBX
ame-MTBY
pig-MTBY
ame-MTAX
ame-MTAY
sha-MTA
pig-MTAX
pig-MTAY
C
Hydrophobic
Polar
Charged
+
-
G T S Q N Y C V L I F W M A P H R K E D
the-MTA3
the-MTA2
the-MTA4
the-MTA5
the-MTA6
the-MTA7
mal-MTA
ame-MTAX
ame-MTAY
sha-MTA
pig-MTAX
pig-MTAY
the-MTB2
the-MTB3
the-MTB6
the-MTB4
the-MTB5
the-MTB7
mal-MTB
bor-MTAX
bor-MTAZ
bor-MTAY
can-MTA
bor-MTBZ
bor-MTX
bor-MTY
can-MTB
ame-MTBX
ame-MTBY
sha-MTB
pig-MTBX
pig-MTBY
D
The-Mal
Pig-Ame
T. sha
Bor-Can
MTA
MTB
PC2 (17.2%)
PC1 (62.8%)
4
2
0
-2
-4
-6
-6
-4
-2
0
2
4
6

Figure S7

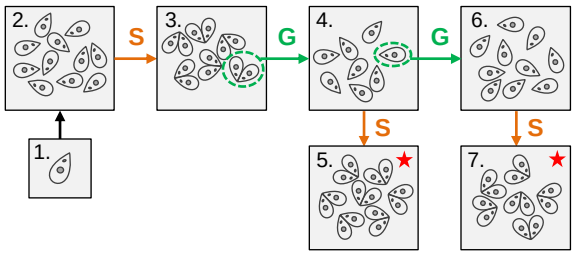


Figure S8

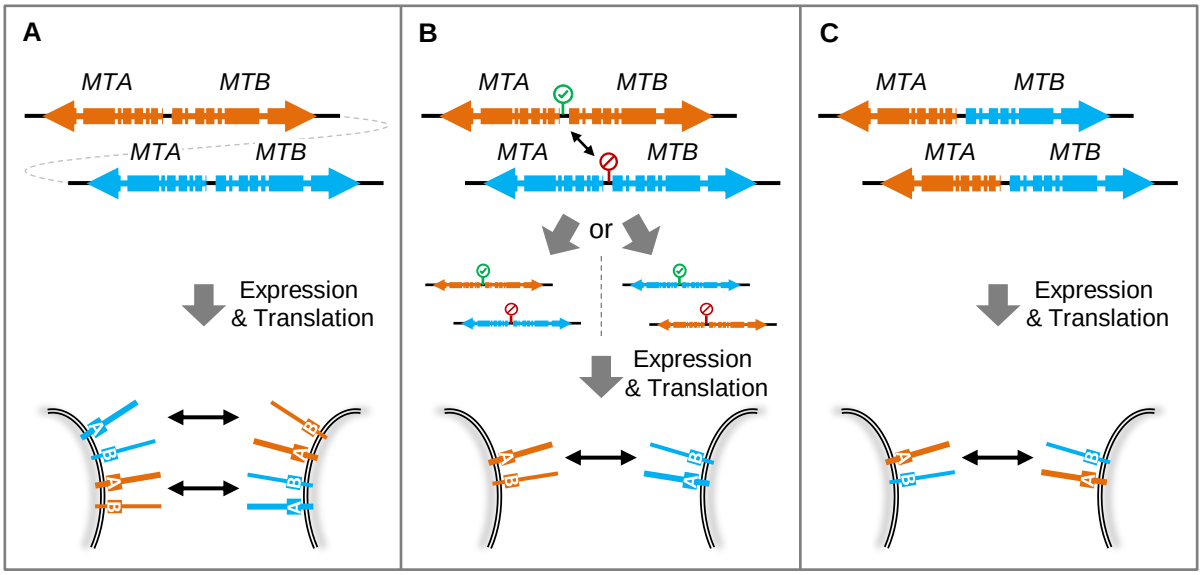

A
MTA
MTB
MTA
MTB
Expression
& Translation
B
MTA
MTB
MTA
MTB
or
Expression
& Translation
C
MTA
MTB
MTA
MTB
Expression
& Translation

Figure S9

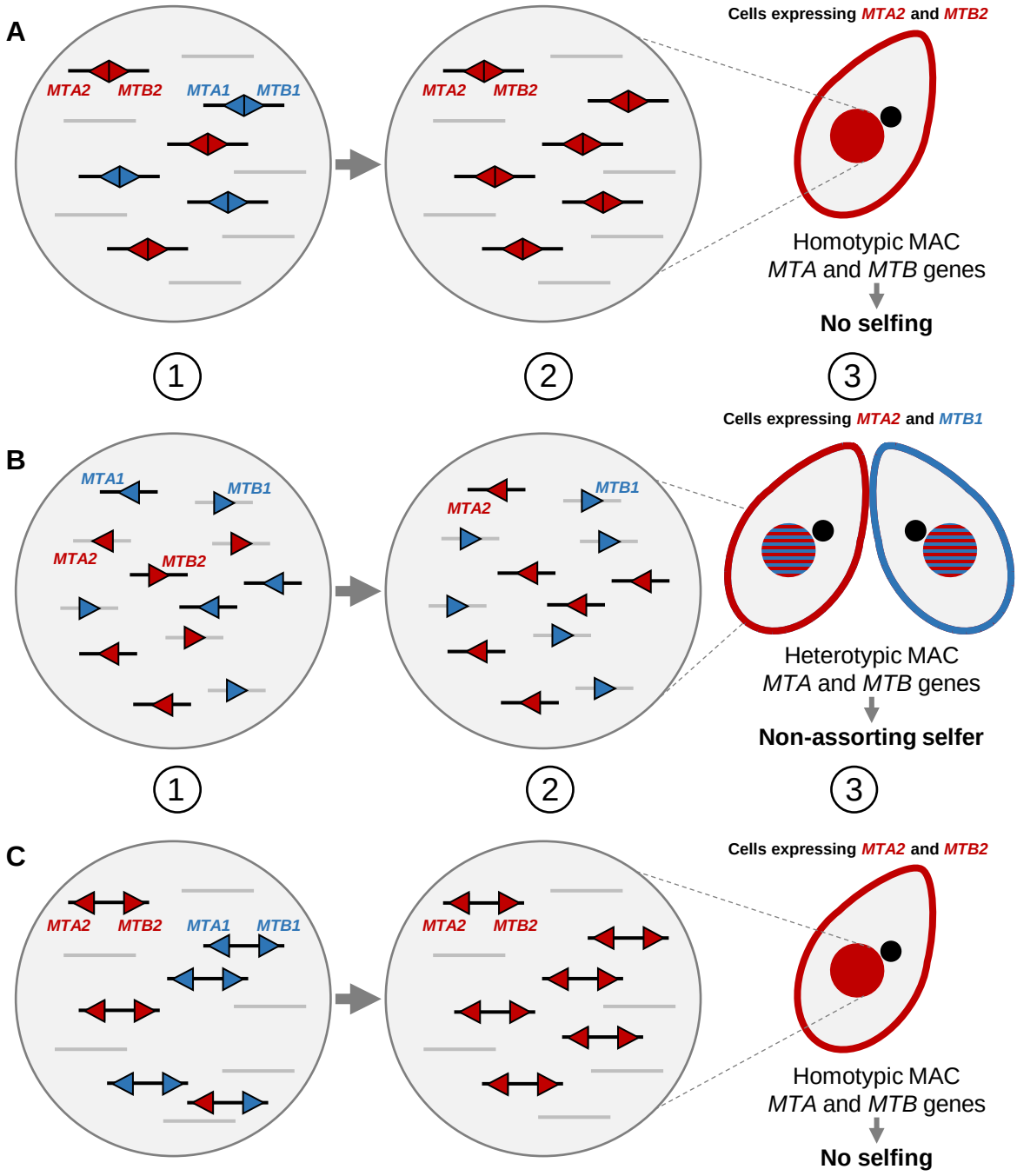

A
MTA2
MTB2
MTA1
MTB1
MTA2
MTB2
Cells expressing MTA2 and MTB2
Homotypic MAC
MTA and MTB genes
No selfing
1
2
3
B
MTA1
MTB1
MTA2
MTB2
MTB1
MTA2
Cells expressing MTA2 and MTB1
Heterotypic MAC
MTA and MTB genes
Non-assorting selfer
1
2
3
C
MTA2
MTB2
MTA1
MTB1
MTA2
MTB2
Cells expressing MTA2 and MTB2
Homotypic MAC
MTA and MTB genes
No selfing

Table S1. Conservation of intron phase and location among functional *Tetrahymena* mating type genes. Related to Figure 3.

Tetrahymena Species	Mating type	MT Gene	Introns and their phases							
			1	2	3	4	5	6	6.5	7
thermophila	II	A	-	1	0	-	1	0	1	0
thermophila	III	A	0	1	0	-	1	0	1	0
thermophila	IV	A	-	1	0	-	1	0	1	0
thermophila	V	A	0	1	0	-	1	0	1	0
thermophila	VI	A	0	1	0	-	1	0	1	0
thermophila	VII	A	0	1	0	-	1	0	1	0
malaccensis	X	A	0	1	0	-	1	0	1	0
borealis	X	A	-	1	0	1	1	0	-	0
borealis	Y	A	0	1	0	0	1	0	1	0
borealis	Z	A	-	2	0	1	1	0	-	0
canadensis	X	A	-	-	0	1	1	0	-	0
shanghaiensis	Selfer	A	0	1	0	1	1	0	-	0
americanis	X	A	0	1	0	1	1	0	-	0
americanis	Y	A	0	1	0	1	1	0	-	0
pigmentosa	I	A	0	1	0	1	1	0	-	0
pigmentosa	III	A	0	1	0	1	1	0	-	0
thermophila	II	B	-	1	0	1	1	0	-	0
thermophila	III	B	-	1	0	1	1	0	-	0
thermophila	IV	B	-	1	0	1	1	0	-	0
thermophila	V	B	-	1	0	1	1	0	-	0
thermophila	VI	B	-	1	0	1	1	0	-	0
thermophila	VII	B	-	1	0	1	1	0	-	0
malaccensis	V	B	-	1	0	1	1	0	-	0
borealis	X	B	-	1	0	1	1	0	-	0
borealis	Y	B	-	1	0	1	1	0	-	0
borealis	Z	B	-	1	0	1	1	0	-	0
canadensis	X	B	-	1	0	1	1	0	-	0
shanghaiensis	Selfer	B	-	1	0	1	1	0	-	0
americanis	X	B	-	1	0	1	1	0	-	0
americanis	Y	B	-	1	0	1	1	0	-	0
pigmentosa	I	B	-	1	0	1	1	0	-	0
pigmentosa	III	B	-	1	0	1	1	0	-	0

Key to cell colors: blue, green and pink: phase 0, 1 and 2 intron, respectively; yellow: missing conserved intron.

Table S2. Number of independent mating tests done within and between sexual *Tetrahymena* strains used. Related to Table 2.

the **mt II**	***the*** **mt III**	***the*** **mt IV**	***the*** **mt V**	***the*** **mt VI**	***the*** **mt VII**	***mal*** **X**	***bor*** **mt X**	***bor*** **mt Y**	***bor*** **mt Z**	***can*** **mt X**	***pig*** **mt I**	***pig*** **mt III**	***ame*** **mt X**	***ame*** **mt Y**	**Mating type**
>>4 neg	>>4 pos	>>4 pos	>>4 pos	>>4 pos	>>4 pos	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	***the*** **mt II**
	>>4 neg	>>4 pos	>>4 pos	>>4 pos	>>4 pos	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	***the*** **mt III**
		>>4 neg	>>4 pos	>>4 pos	>>4 pos	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	***the*** **mt IV**
			>>4 neg	>>4 pos	>>4 pos	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	***the*** **mt V**
				>>4 neg	>>4 pos	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	***the*** **mt VI**
					>>4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	***the*** **mt VII**
						4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	***mal*** **mt X**
							4 neg	4 pos	4 pos	4 neg	4 neg	4 neg	4 neg	4 neg	***bor*** **mt X**
								4 neg	4 pos	4 neg	4 neg	4 neg	4 neg	4 neg	***bor*** **mt Y**
									4 neg	4 neg	4 neg	4 neg	4 neg	4 neg	***bor*** **mt Z**
										4 neg	4 neg	4 neg	4 neg	4 neg	***can*** **mt X**
											>4 neg	>4 pos	4 neg	4 neg	***pig*** **mt I**
												>4 neg	4 neg	4 neg	***pig*** **mt III**
													>4 neg	>4 pos	***ame*** **mt X**
														>3 neg	***ame*** **mt Y**

Key to entries:

the	*T. thermophila*
mal	*T. malaccensis*
bor	*T. borealis*
can	*T. canadensis*
pig	*T. pigmentosa*
ame	*T. americanis*
pos	Mating observed
neg	No mating observed
>> 4	Many independent tests, including for many experiments unrelated to this project.
>4	More that 4 independent tests, including for experiments unrelated to this project.

Table S3. Collection sites and other information on previously unreported strains. Related to Table 2.

Species	Mating type	Strain	Collection site	Type	State	Latitude	Longitude
T. americanis	X	SD21194	LakeBarkley1	lake	KY	37.05093	-88.15287
T. americanis	Y	SD21244	ShantyHollowLake	lake	KY	37.14453	-86.38432
T. borealis	Y	SD19502	KinzuaBayElijah	lake	PA	41.81892	-78.94685
T. borealis	Z	SD19803	LittleSalmonCr165	stream	PA	41.50955	-79.15158
T. pigmentosa	III	SD19481	FishCrSloughAshland	stream	WI	46.58583	-90.932
T. pigmentosa	I	SD20427	HalfMoonPondBrook	stream	NH	43.17025	-72.08968

Data S1. Evolution of amino acid usage among the mating type genes of sexual *Tetrahymena* species.

In main text we have reported evidence of co-variation in *MTA* and *MTB* genes that could reflect changes in putative requirements for functional protein interactions between the two gene products (Orias et al., 2017). To take a deeper look, we examined amino acid usage in the MTA and MTB proteins of sexual species. Hydrophobicity, for example, can influence protein flexibility during protein-protein interaction (Radivojac et al., 2004). By using either the hierarchical or k-means method, we clustered the amino acid usage frequencies of the MTA and MTB proteins separately for a) the entire lengths of the proteins, b) the proximal ~2/3 of the proteins (from N-terminus to intron 7), and c) the distal ~1/3 (C-terminal exon) of the proteins. The results are shown in the 3 panels of **Figure S6**. Comparison of the three clustering trees shows that they are different from one another and that none of them have the same topology as either the species phylogenetic tree (**Figure 2C**) or the mating type protein trees for each of the three segments (**Figure 4**).

The following unusual topologies were common to all three amino acid usage clustering trees.

1) The main branches of the amino acid clustering trees tend to be different species phylogenetic groups rather the MTA vs. MTB main branching found in the protein phylogenetic tree.

2) Amino acid usage the "Bor-Can" subclade of the borealis clade clusters with the "Australis" clade rather than with the other ("The-Mal") subgroup of the "Borealis" clade.

Although we don't have a clear rationale for these differences, we can understand some of the variables influencing the branching pattern of particular amino acid usage clustering trees for the distal 1/3 of the MTA and MTB proteins.

a) For *T. thermophila*: the terminal half of the terminal exon (distal ~1/6) of both *MTA* and *MTB* genes are identical because the tandem array of MIC mtGPs contain only one copy of the *MTA* distal ~1/6 (at the "head" of the array), and only one copy of the *MTA* distal ~1/6 (at the "tail" of the array). Even for the rest of the C-terminal exon, the copies present in the six genes are >95% identical to one another (Cervantes et al., 2013). Thus, within the *T. thermophila* branch of that distal 1/3 clustering tree, all the MTA proteins must cluster together in one subcluster, while all the MTB proteins must cluster together in the other subcluster, as they indeed do.

b) For the "Australis" clade: although the mtGP composition of their MIC mating type locus has not yet been described, we show here that the two sequenced MTA proteins in each species investigated (*pigmentosa* and *americanis*) show high sequence conservation with one another. The same is true for MTB proteins. Interestingly, for the "Australis" clade, the high degree of amino acid sequence conservation seen in the distal ~1/3 extends to the "middle third". This has to influence the branching pattern seen for the proximal 2/3 clustering tree in a different way for "Australis" species than for the rest of the species.

c) For *T. borealis*, there is essentially no sequence conservation in the distal 1/3 of either MTA or MTB proteins. This disorganizes the branching patterns seen for the three MTAs and MTB proteins relative to the other groups.

A "Principal Components Analysis" (PCA) of the amino acid usage frequency for each full-length mating type protein of every *Tetrahymena* sexual species (**Figure S6D**) shows a similar clustering pattern as that obtained by either the hierarchical or k-means method of the full-length proteins (**Figure S6A**).

Although we can make some sense of the amino acid usage clustering we observe, it is clear that to fully understand the details of the clustering we must wait until we have an understanding of the 3-D structure of these proteins and experimental analyses of their interactions.

References

CERVANTES, M. D., HAMILTON, E. P., XIONG, J., LAWSON, M. J., YUAN, D., HADJITHOMAS, M., MIAO, W. & ORIAS, E. 2013. Selecting one of several mating types through gene segment joining and deletion in *Tetrahymena thermophila*. *PLoS Biol,* 11**,** e1001518.

ORIAS, E., SINGH, D. P. & MEYER, E. 2017. Genetics and Epigenetics of Mating Type Determination in *Paramecium* and *Tetrahymena*. *Annual Review of Microbiology,* 71.

RADIVOJAC, P., OBRADOVIC, Z., SMITH, D. K., ZHU, G., VUCETIC, S., BROWN, C. J., LAWSON, J. D. & DUNKER, A. K. 2004. Protein flexibility and intrinsic disorder. *Protein Science,* 13**,** 71–80.

Data S2. Classification of *Tetrahymena* selfers.

While trying to distinguish various types of selfers encountered in this work, in the context of our current molecular knowledge about *Tetrahymena* mating types, it became useful to devise a more descriptive categorization of *Tetrahymena* selfers. Our proposed classification, summarized in **Table SD1** below, is based on a combination of the genetic behavior of selfers during asexual reproduction and after sexual reorganization upon selfing. We consider it a descriptive classification because more than one molecular mechanism may be found to generate the same type of selfer genetic behavior.

A. Selfing behavior during asexual reproduction.

1) Assorting selfers. When sexual progeny become sexually mature, a mixture of mating type alleles is often found in the MAC; if allowed to starve, subcultures of these cell lines will self. Upon further asexual propagation, the mating type alleles ultimately assort, so cells become pure for a single MAC mating type allele. At this point, they, and their subsequent vegetative descendants no longer self and instead express a defined mating type. Examples of such selfers are the "classical" selfers analyzed in *T. thermophila* by Allen and Nanney (1958).

2) Non-assorting selfers. Upon reaching sexual maturity, progeny cells and their vegetative descendants always self upon starvation. No vegetative assortants expressing a single mating type are obtained. Examples are the "suicidal" *Tetrahymena* selfers described by Nanney (1953), the persistent selfers reported by Lin and Yao (2020) and the *T. shanghaiensis* perpetual selfers investigated in this study.

B. Perpetuation of selfing after sexual reorganization and maturity.

Two possibilities are known:

1) Sexually non-true-breeding selfers. The sexual progeny of selfing pairs are not themselves selfers when they reach sexual maturity. Examples are the selfers investigated by Nanney and Allen listed under **A1** above, where the old MAC, responsible for selfing, is destroyed upon conjugation. Persistent selfers obtained by post-conjugation Ku80 silencing (Lin and Yao, 2020) should also fall in this category.

2) Sexually true-breeding selfers. The sexual progeny of non-assorting selfers are themselves always vegetatively non-assorting selfers. The *T. shanghaiensis* selfers used in this

study are an example.

Based on the above classifications, the ordinary *T. thermophila* selfers analyzed by Allen and Nanney (1958), are classified as vegetatively assorting, sexually non-true-breeding selfers. In contrast, the *T. shanghaiensis* strain used here is a vegetatively non-assorting, sexually true-breeding selfer.

Other types of selfers may well be discovered in nature or in the lab.

Table SD1. Proposed classification and examples *Tetrahymena* selfers. Related to Figure S7.

Sexual Behavior	**Asexual Behavior**	
	Vegetatively assorting	**Vegetatively non-assorting**
Sexually non-true-breeding	*T. thermophila* "classical" selfers (Allen and Nanney, 1958)	
Sexually true-breeding		*T. shanghaiensis* selfers (this article) and newly described *T. thermophila* persistent selfers (Lin and Yao, 2020)

References

ALLEN, S. L. & NANNEY, D. L. 1958. An Analysis of Nuclear Differentiation in the Selfers of *Tetrahymena*. *The American Naturalist,* 92**,** 139-160.

LIN, I. T. & YAO, M. C. 2020. Selfing mutants link Ku proteins to mating type determination in *Tetrahymena*. *PLoS Biol,* 18**,** e3000756.

NANNEY, D. L. 1953. Nucleo-cytoplasmic interaction during conjugation in *Tetrahymena*. *The Biological Bulletin,* 105**,** 133-148.

Data S3. Evolution of *MTA-MTB* contiguity after functional differentiation would have discouraged selfing.

The putative events that led to the evolution of the *Tetrahymena* mtGP have been presented (**Figure 6**) and discussed in main text. We have argued that *MTA* and *MTB* were not tightly linked to one another at the time they evolved their distinct and complementary functions in co-stimulation and pair formation. Otherwise, frequent gene conversion within the palindromic duplication would have acted to maintain the similarity of the two genes and would have precluded the evolution of their functional differences.

The possible non-adjacency of *MTA* and *MTB*, once the two genes had evolved functional differences and replaced an older mating type system, is interesting with regard to selfing. The frequency of selfing after allelic assortment is complete would have depended on the location of the two genes in a common ancestor of the "Borealis" and "Australis" clades, as illustrated in **Figure S9**. If *MTA* and *MTB* resided on different MAC chromosomes when they evolved their functional differences, the probability of generating MAC genotypes capable of resulting in non-assorting selfers would have been high (**Figure S9B**) compared to what we see now (**Figure S9A**).

The opposite is true if *MTA* and *MTB* had evolved at locations that ended up on the same MAC chromosome **Figure S9, panel C**). In this case, the probability of selfing would be very low because intra-chromosomal MAC recombination is rare between genes on the same MAC chromosome (Longcor et al., 1996), likely related to infrequent DNA damage repair events occurring between the two genes during asexual reproduction. The heterotypic MAC chromosome shown at the bottom of **Figure S9, panel C1** is an example of such rare MAC intrachromosomal recombination. If the actual number of chromosome types in such MAC had been 22 *MTA1-MTB1* copies, 22 *MTA2-MTB2* copies and 1 *MTA1-MTB2* copy, then, after complete assortment of this clone, the percentage of assortants would have been 49% mt I, 49% mt II and 2% non-assorting selfers, respectively.

Recombination events (MIC or MAC) between *MTA* and *MTB* genes in the current mtGPs are extremely rare. No cases have been reported in the extensive laboratory work with *T. thermophila*, but such an event is a possible basis for the naturally occurring heterotypic *T. shanghaiensis* mtGP described in this article. The low probability of recombination is explained by the main features of the contemporary mtGP: the side by side location of the genes and the short length and mating type-specific sequence diversification of the intergenic region.

Reference

LONGCOR, M. A., WICKERT, S. A., CHAU, M.-F. & ORIAS, E. 1996. Coassortment of genetic loci during macronuclear division in *Tetrahymena thermophila*. *European Journal of Protistology,* 32**,** 85-89.

Transparent Methods

Biological methods

The *Tetrahymena* species whose genomes we sequenced were identical to those in our recent comparative genomics report (Xiong et al., 2019). *T. americanis* and *T. pigmentosa* strains were provided by Dr. Paul Doerder (Cleveland State University, USA). All cells were grown in SPP medium (1% Proteose Peptone, 0.2% glucose, 0.1% yeast extract and 0.003% Sequestrene).

Mating type differences between strains of the same species were tested by a pairing assay. To induce starvation, cells were washed twice with 10 mM Tris-Cl (pH 7.4) and then resuspended in the same solution and incubated for at least 12 h, at which this time no obvious food vacuoles could be observed. For the mating tests, equal numbers of fully starved cells of two different strains were mixed at a final cell density of approximately 2×10^5 cells/ml (final volume ~1 ml). If any pairs formed (usually over 80%), and no pairs were found in the starved, unmixed controls, then the two strains were considered to be of different mating type. Every mating test was repeated at least four times (see **Table S2** for more details).

Identification of mating type homologs in published *Tetrahymena* genomes

A total of ten *Tetrahymena* genomes were chosen for our analyses: *T. thermophila*, *T. malaccensis*, *T. pyriformis*, *T. vorax*, *T. borealis*, *T. canadensis*, *T. pigmentosa, T. americanis, T. shanghaiensis*, and *T. paravorax* (**Table 1**).Candidate mating type gene homologs in genome sequences were determined by performing a local tBLASTn (version 2.2.25) (Altschul et al., 1997) search against the *T. thermophila MTA* and *MTB* genes (Cervantes et al., 2013). To confirm intron/exon structures, RNA-Seq data was mapped to each genome with TopHat (version 2.0.9) and visually checked using IGV (version 2.4.6). This resulted in a correction of previously predicted (Xiong et al., 2019) intron/exon structures for *T. canadensis* and *T. shanghaiensis*. To check whether mating type genes were at the same MAC genome location as in *T. thermophila*, chromosome synteny analysis of the surrounding regions was done based on the *Tetrahymena* comparative genomics database (Yang et al., 2019) (http://ciliate.ihb.ac.cn/).

Identification of mating type homologs in strains without sequenced genomes

Complete mating type homologs, in strains without sequenced genomes, were first identified by RNA sequencing; PCR was used to fill any gaps. This method was used for two strains of *T. borealis*, *T. americanis*, and *T. pigmentosa* (**Table 1**). Strain details are shown in **Table S3**. Total RNA was extracted from each strain at the starvation stage using the RNeasy Protect Cell Mini Kit (Qiagen), as described (*Tetra*FGD) (Xiong et al., 2011). Poly-A tailed mRNA was then enriched using Sera-Mag magnetic oligo (dT) beads. Illumina sequencing libraries were constructed and paired-end (150 bp × 2) sequencing was done for all samples using Illumina Hiseq4000 sequencer. After adaptor trimming by Trim-Galore (version 0.4.0) (Wu et al., 2011), clean reads were *de novo* assembled using Trinity (version trinityrnaseq_r20140717) (Grabherr et al., 2011). The resulting transcriptomes were subsequently searched with *T. thermophila MTA* and *MTB* gene sequence using tBLASTn (version 2.2.25). Usually only incomplete transcripts were assembled; PCR amplification and Sanger sequencing were used to link these fragments and obtain the intergenic regions of the mating type genes. PCR primers are listed in **Table S4**. To confirm the intron/exon structures, clean reads were mapped back to the sequence with TopHat (version 2.0.9) (Kim and Salzberg, 2011) and visually checked by IGV (Integrative Genome Viewer; version 2.4.6) (Robinson et al., 2011, Thorvaldsdóttir et al., 2013).

The sequences of MAC mating type loci generated in this study have been deposited in GenBank (**Table S5**). Genome data for all strains except *T. pigmentosa* and *T. americanis* (which do not yet have complete genome assemblies) can be accessed through *Tetrahymena* comparative genomics database (http://ciliate.ihb.ac.cn/); additional sequencing data has been submitted under accession numbers PRJNA510545.

Estimation of gene expression levels

To estimate gene expression levels, RNA sequencing reads were mapped to the MAC mating type loci with TopHat (version 2.0.9) and the FPKM value was obtained using Cufflinks (version 2.1.1) (Trapnell et al., 2012).

Sequence analyses

To calculate GC content, MAC Mating type loci were cut every 50 bp using a sliding-window approach. Multiple sequence alignment was done using ClustalW (Larkin et al., 2007) and the results were visualized in ESPrint 3 (Robert and Gouet, 2014). Positional sequence

similarity was calculated based on the alignment results and were averaged using a 20 aa sliding-window. To construct phylogenetic tree, protein sequences were aligned by ClustalW, and sites with over 50% gaps were deleted. Then the best-fit model was calculated by ProtTest (version 3.4.2) (Darriba et al., 2011), and the phylogenetic tree was generated by IQ-TREE (version 1.6.12) (Nguyen et al., 2015). Principal Components Analysis and protein clustering according to the percentage of amino acids were done by MeV (version 4.9.0) (Euclidean distance was used) (Saeed et al., 2003). ANOVA and t-test were done using GraphPad (version 8.0.2).

References

ALTSCHUL, S. F., MADDEN, T. L., SCH FFER, A. A., ZHANG, J., ., ZHANG, Z., ., MILLER, W., . & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* 25**,** 3389-3402.

CERVANTES, M. D., HAMILTON, E. P., XIONG, J., LAWSON, M. J., YUAN, D., HADJITHOMAS, M., MIAO, W. & ORIAS, E. 2013. Selecting one of several mating types through gene segment joining and deletion in *Tetrahymena thermophila*. *PLoS Biol,* 11**,** e1001518.

DARRIBA, D., TABOADA, G. L., DOALLO, R. & POSADA, D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics,* 27**,** 1164-1165.

GRABHERR, M. G., HAAS, B. J., YASSOUR, M., LEVIN, J. Z., THOMPSON, D. A., AMIT, I., ADICONIS, X., FAN, L., RAYCHOWDHURY, R., ZENG, Q., CHEN, Z., MAUCELI, E., HACOHEN, N., GNIRKE, A., RHIND, N., DI PALMA, F., BIRREN, B. W., NUSBAUM, C., LINDBLAD-TOH, K., FRIEDMAN, N. & REGEV, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol,* 29**,** 644-52.

KIM, D. & SALZBERG, S. L. 2011. TopHat-Fusion: An algorithm for discovery of novel fusion transcripts. *Genome Biol,* 12**,** R72.

LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A., LOPEZ, R., THOMPSON, J. D., GIBSON, T. J. & HIGGINS, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23**,** 2947-8.

NGUYEN, L. T., SCHMIDT, H. A., VON HAESELER, A. & MINH, B. Q. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol,* 32**,** 268-74.

ROBERT, X. & GOUET, P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res,* 42**,** W320-4.

ROBINSON, J. T., THORVALDSDOTTIR, H., WINCKLER, W., GUTTMAN, M., LANDER, E. S., GETZ, G. & MESIROV, J. P. 2011. Integrative genomics viewer. *Nat Biotechnol,* 29**,** 24-6.

SAEED, A. I., SHAROV, V., WHITE, J., LI, J., LIANG, W., BHAGABATI, N., BRAISTED, J., KLAPA, M., CURRIER, T., THIAGARAJAN, M., STURN, A., SNUFFIN, M., REZANTSEV, A., POPOV, D., RYLTSOV, A., KOSTUKOVICH, E., BORISOVSKY, I., LIU, Z., VINSAVICH, A., TRUSH, V. & QUACKENBUSH, J. 2003. TM4: a free, open-source system for microarray

data management and analysis. *Biotechniques,* 34**,** 374-8.

THORVALDSD TTIR, H., ROBINSON, J. T. & MESIROV, J. P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics,* 14**,** 178-192.

TRAPNELL, C., ROBERTS, A., GOFF, L., PERTEA, G., KIM, D., KELLEY, D. R., PIMENTEL, H., SALZBERG, S. L., RINN, J. L. & PACHTER, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc,* 7**,** 562-78.

WU, Z., WANG, X. & ZHANG, X. 2011. Using non-uniform read distribution models to improve isoform expression inference in RNA-Seq. *Bioinformatics,* 27**,** 502-8.

XIONG, J., LU, X., LU, Y., ZENG, H., YUAN, D., FENG, L., CHANG, Y., BOWEN, J., GOROVSKY, M., FU, C. & MIAO, W. 2011. *Tetrahymena* Gene Expression Database (TGED): a resource of microarray data and co-expression analyses for *Tetrahymena*. *Sci China Life Sci,* 54**,** 65-7.

XIONG, J., YANG, W., CHEN, K., JIANG, C., MA, Y., CHAI, X., YAN, G., WANG, G., YUAN, D., LIU, Y., BIDWELL, S. L., ZAFAR, N., HADJITHOMAS, M., KRISHNAKUMAR, V., COYNE, R. S., ORIAS, E. & MIAO, W. 2019. Hidden genomic evolution in a morphospecies—The landscape of rapidly evolving genes in *Tetrahymena*. *PLOS Biology,* 17**,** e3000294.

YANG, W., JIANG, C., ZHU, Y., CHEN, K., WANG, G., YUAN, D., MIAO, W. & XIONG, J. 2019. Tetrahymena Comparative Genomics Database (TCGD): a community resource for *Tetrahymena*. *Database (Oxford),* 2019.

Table SE1. Primers used to fill introns and intergenic region sequence.

Species	Mating type	Primer	Sequence (5′→3′)
T. borealis	Y	BorII-1183-F	TTACTAGCAACCAGAGCATTGATG
		BorII-3109-R	AACTTAATAGCCAAATTTATTAAAG
		BorII-2889-F	TAATAGGTATAACTGGAGGTGGTG
		BorII-2889-Fs*	TGGCAAATATCTCCAGATTAAGA
		BorII-4346-F	TTAATAGGTTACCCATAAGCCG
		BorII-4940-Rs*	TAATCTTTCCAAATCGGATTTT
		BorII-4940-R	CATCAGAAGAATCTCGTGAAGATA
		BorII-6393-Rs*	GTAGAAGCAAGTGAAGATGGATAAG
		BorII-6393-R	AGAGTTAAGAGGAATAGCAGGAAG
		BorII-6217-F	TATTTGATGCTTTTTCTCTAGATG
		BorII-6217Fs*	GTATCTATTCTTAAAGGATCAA
		BorII-8515-R	AATAGTTTAGGCTGCATTGAGTAC
		BorII-9760-R	TAGCAGCTAAATAAACAGATAGAAC
	Z	BorIII-1201-F	CATTGCATTGCTTACATTCTTAG
		BorIII-3289-R	CAAATGTCTTATGATGAAAGGATTG
		BorIII-3027-F	TTTGTATTCCAATCTTTTAGTCACG
		BorIII-4846-R	GAAGAAATCCCGAATTATGAAAG
		BorIII-4332-F	ATTTAGGTATCGCTTTATGTGAATG
		BorIII-6428-R	TTAGAGTTTAAGTACTGAGGCTAAC
		BorIII-5500-F	TGTTACCCCTTCAGGTTATTTTC
		BorIII-5500-Fs*	TTTGATTAAGTATGCTTGCCCT
		BorIII-7565-F	ATAGATAATTCTGCTCTAGCCTTAG
		BorIII-9070Rs*	TACATTCCTTTTATTCGCATTC
		BorIII-9070R	TGCCCTTAGTCTATTGGTTGC
T. pigmentosa	I	Pig-1325-F	TATCTCCCACTACATCCAGAACAAG
		Pig-3194-R	CATCTGTCTCATGTTACAATCAAGC
		Pig-2861-F	TCAATGCTCGTCATATTTAACTG
		Pig-2861-Fs*	ACCTGATAATTCGATTGACT
		PigI-5069-Rs*	TCCTCTTTTCAAAACAGCAA
		PigI-5069-R	GGAATATGAATGCAAATACGAG
		PigI-4572-F	CTTGAATTTAATTGCTGTTTTG
		PigI-4572-Fs1*	CTCAATAGATGAGCTCTCCATA
		PigI-4572-Fs2*	ACTATGAGATTACCTTCGATGT
		PigI-4572-Fs3*	AGTTTATGCTTTCCTTACGAGATG
		Pig-6984-F	TCCCTGACAATTATCCGAAAC
		Pig-8757-R	TGTAGGATTGGGAGGGTTTG
	III	Pig-1325-F	TATCTCCCACTACATCCAGAACAAG
		Pig-3194-R	ATCTGTCTCATGTTACAATCAAGC
		Pig-2861-F	TCAATGCTCGTCATATTTAACTG
		Pig-2861-Fs*	CTCATTACCTCCAGGAAAACTTAT

		PigIII-4214-F	TTCACCAATATTTAGTTATTAGCTG
		PigIII-4214-Fs*	ACTCAAATGAAGGATATGAAAAC
		PigIII-6856-R	TAAATGCTTGAGGAATAGTGAATG
		PigIII-5750-F	AATCTGCCCTAAGCAATGC
		PigIII-6986-R	GGTTTCGGATAATTGTCAGG
		Pig-6984-F	TCCCTGACAATTATCCGAAAC
		Pig-8757-R	TGTAGGATTGGGAGGGTTTG
T. americanis	X	Ame-749-F	CCTAATGCTCCGATTTCTGC
		Ame-2196-R	GAAGTTATCCTGCCACTCAAAC
		Ame-2074-F	GATGATTAAGTGAACGTACTCAAGA
		AmeI-3341-R	GCCATGGTGTTGTTTCTAGG
		AmeI-3136-F	ATTTCCTCCACTTAAATCAGATAT
		AmeI-3136-Fs*	CTCTAACTAAAACTTGGACTATCTT
		AmeI-4413-R	TATGCATTAAACTAAGAGGGTGT
		AmeI-4072-F	ACTACATTCTACACAATCCATCTTA
		AmeI-5013-R	TTCCTGATTTCAGCATTGG
		AmeI-4877-F	CTATGGATGTATGAGCACTGCT
		Ame-5968-R	TTTATTTACAACAGATGGAGCAG
		AmeI-5522-F	AGACCTGATTGCCATAGACTATAG
		Ame-6475-R	AGATTGAGCTGGCTTAGGGT
		Ame-6358-F	AAGCATAGCCTCTGAAGATAAA
		Ame-8111-R	GCATTATTTATTTATGGCATCTAC
	Y	Ame-749-F	CCTAATGCTCCGATTTCTGC
		Ame-2196-R	GAAGTTATCCTGCCACTCAAAC
		Ame-2074-F	GATGATTAAGTGAACGTACTCAAGA
		AmeII-4085-Rs*	TGAAGGCCACAGTAATAAGAAAT
		AmeII-4085-R	GTGTTGAATGCACCCTCTAT
		AmeII-3767-F	TCGAAATCTTCCTTCCATACAG
		AmeII-3767-Fs*	CCCATTACATTCTAATCTAAGTCG
		AmeII-4887-R	TTAAAGAAGTGCTCATACACCC
		AmeII-4723-F	TGAGAATAAGAGATGATGGCAG
		Ame-5968-R	TTTATTTACAACAGATGGAGCAG
		AmeII-5707-F	AATTAAAGCTTATTTGTAACGTGTT
		Ame-6475-R	AGATTGAGCTGGCTTAGGGT
		Ame-6358-F	AAGCATAGCCTCTGAAGATAAA
		Ame-8111-R	GCATTATTTATTTATGGCATCTAC

* Primers only used for sequencing.

Table SE2. Accession number of *Tetrahymena* mating type genes.

Species	Strain	Mating type	Sequence source	Accession number	Reference (s)
T. thermophila	SB4208	II	PCR & Transcriptome	KC405255	Cervantes et al, 2013
	SB4213	III	PCR & Transcriptome	KC405261	Cervantes et al, 2013
	SB4214	IV	PCR & Transcriptome	KC405259	Cervantes et al, 2013
	SB4218	V	PCR & Transcriptome	KC405256	Cervantes et al, 2013
	SB4220	VI	Genome	KC405258	Eisen et al, 2006; Cervantes et al, 2013
	SB4223	VII	PCR & Transcriptome	KC405260	Cervantes et al, 2013
T. malaccensis	SD01608	-	Genome	MK315120	Xiong *et al*, 2019
T. pyriformis	GL	-	Genome	MK315121	Xiong *et al*, 2019
T. vorax	SD30421	-	Genome	MK315122	Xiong *et al*, 2019
T. borealis	SD01609	X	Genome	MK315123	Xiong *et al*, 2019
	SD19502	Y	Transcriptome & PCR	MK315124	This report
	SD19803	Z	Transcriptome & PCR	MK315125	This report
T. canadensis	SD30770	-	Genome	MK315126	Xiong *et al*, 2019
T. shanghaiensis	SD205039	-	Genome	MK315127	Xiong *et al*, 2019
T. pigmentosa	SD19481	X	Transcriptome & PCR	MK315128	This report
	SD20427	Y	Transcriptome & PCR	MK315129	This report
T. americanis	SD21194	X	Transcriptome & PCR	MK315130	This report
	SD21244	Y	Transcriptome & PCR	MK315131	This report
T. paravorax	SD205177	-	Genome	MK315132	Xiong *et al*, 2019