ARTICLE

doi:10.1038/nature13318

Genome-defence small RNAs exapted for epigenetic mating-type inheritance

Deepankar Pratap Singh[1,2], Baptiste Saudemont[1,2]†, Gérard Guglielmi[1], Olivier Arnaiz[3], Jean-François Goût[4]†, Malgorzata Prajer[5], Alexey Potekhin[6], Ewa Przybòs[5], Anne Aubusson-Fleury[3], Simran Bhullar[1], Khaled Bouhouche[1]†, Maoussi Lhuillier-Akakpo[2,7], Véronique Tanty[1], Corinne Blugeon[1], Adriana Alberti[8], Karine Labadie[8], Jean-Marc Aury[8], Linda Sperling[3], Sandra Duharcourt[7] & Eric Meyer[1]

In the ciliate *Paramecium*, transposable elements and their single-copy remnants are deleted during the development of somatic macronuclei from germline micronuclei, at each sexual generation. Deletions are targeted by scnRNAs, small RNAs produced from the germ line during meiosis that first scan the maternal macronuclear genome to identify missing sequences, and then allow the zygotic macronucleus to reproduce the same deletions. Here we show that this process accounts for the maternal inheritance of mating types in *Paramecium tetraurelia*, a long-standing problem in epigenetics. Mating type E depends on expression of the transmembrane protein mtA, and the default type O is determined during development by scnRNA-dependent excision of the *mtA* promoter. In the sibling species *Paramecium septaurelia*, mating type O is determined by coding-sequence deletions in a different gene, *mtB*, which is specifically required for *mtA* expression. These independently evolved mechanisms suggest frequent exaptation of the scnRNA pathway to regulate cellular genes and mediate transgenerational epigenetic inheritance of essential phenotypic polymorphisms.

Ciliates are complex unicellular eukaryotes that use different types of nuclei within the same cytoplasm to separate germline and somatic functions[1]. The diploid micronuclei (MICs) undergo meiosis to provide gametic nuclei during sexual events, but their genome is not expressed. Genes are expressed from the polyploid macronucleus (MAC), which is not transmitted across sexual generations. After meiosis and fertilization, the parental MAC is lost and replaced by a new one that develops from a mitotic copy of the zygotic nucleus. Macronuclear development involves extensive rearrangements of the germline genome, including the elimination of virtually all transposable elements and other repeats[2]. Furthermore, in *P. tetraurelia* ~45,000 short, single-copy internal eliminated sequences (IESs) are precisely excised from coding and non-coding sequences[3,4] by the domesticated transposase Pgm[5]. IESs are invariably flanked by two 5′-TA-3′ dinucleotides which recombine into one after excision[4]. A short consensus adjacent to the TAs (5′-TAYAGYNR-3′) is reminiscent of the ends of Tc1/mariner elements[4,6], and a recent study provided support for the hypothesis that IESs are degenerate remnants of ancient transposable element insertions[3,7].

All intragenic IESs identified so far must be excised to reconstitute functional genes in the MAC. However, the poorly conserved IES end consensus is not sufficient to specify the excision pattern genome-wide[4], and a specific class of small RNAs is required to identify some IESs on the basis of their absence from the parental MAC[8–10]. In the current 'genome scanning' model, scnRNAs are produced from most of the germline genome during MIC meiosis[11,12] and are then filtered by pairing interactions with nascent transcripts in the parental MAC[13], which acts as a sponge to remove matching scnRNAs from the active pool. Those that cannot find a match remain free to target homologous sequences in the zygotic MAC when it develops, thereby recruiting the IES excision machinery. This RNA-mediated genomic subtraction thus reproduces the deletions observed in the parental MAC and can account for epigenetic inheritance of alternative rearrangement patterns, such as retention of a given IES in the MAC[14,15], or deletion of a given gene[16], across sexual generations.

In *Paramecium*, conjugation (the reciprocal fertilization of cells of opposite mating types) does not allow any significant exchange of cytoplasm between the mates, so that the pools of scnRNAs produced during meiosis are independently sorted in each cell. After fertilization, each developing MAC will thus reproduce the particular rearrangements present in the old MAC of its own cytoplasmic parent. This mechanism might underlie the maternal (cytoplasmic) inheritance of mating types in *P. tetraurelia*, one of the earliest cases of transgenerational epigenetic inheritance in any eukaryote.

Although mating types were discovered in 1937 (ref. 17), so far the only available test relies on the ability of type O (odd) to agglutinate with type E (even) when vegetative cells become sexually reactive, a physiological state induced by mild starvation. Agglutination is a prerequisite for conjugation and occurs on contact through adhesion of ciliary membranes. Mutational analyses showed that several genes are specifically required for expression of type E[18–20]; mutations in these genes result in a constitutive O phenotype, which thus seems to be a default state. In *P. tetraurelia*, mating types are not genetically determined in the MIC[21]. Each new MAC becomes determined for one type during its development, and remains the same throughout vegetative growth of the derived clone. The O/E alternative is not random, but maternally inherited; experiments showed that mating-type determination in the developing zygotic MAC is controlled through the cytoplasm[22] by the maternal MAC[23,24]. A pleiotropic mutation enforcing constitutive determination for type E was later found to impair a limited subset of genome rearrangements[25], suggesting that type O is normally

[1]Ecole Normale Supérieure, Institut de Biologie de l'ENS, IBENS; Inserm, U1024; CNRS, UMR 8197 Paris F-75005, France. [2]Sorbonne Universités, UPMC Univ., IFD, 4 place Jussieu, 75252 Paris cedex 05, France. [3]CNRS UPR3404 Centre de Génétique Moléculaire, Gif-sur-Yvette F-91198, and Université Paris-Sud, Département de Biologie, Orsay F-91405, France. [4]CNRS UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Université de Lyon, 43 boulevard du 11 Novembre 1918, Villeurbanne F-69622, France. [5]Institute of Systematics and Evolution of Animals, Polish Academy of Sciences, Sławkowska 17, 31-016 Krakow, Poland. [6]Department of Microbiology, Faculty of Biology, St Petersburg State University, Saint Petersburg 199034, Russia. [7]Institut Jacques Monod, CNRS, UMR 7592, Université Paris Diderot, Sorbonne Paris Cité, Paris F-75205, France. [8]Commissariat à l'Energie Atomique (CEA), Institut de Génomique (IG), Genoscope, 2 rue Gaston Crémieux, BP5706, 91057 Evry, France. †Present addresses: Laboratoire de Biochimie, Unité Mixte de Recherche 8231, École Supérieure de Physique et de Chimie Industrielles, 75231 Paris, France (B.S.); Department of Biology, Indiana University, Bloomington, Indiana 47405, USA (J.-F.G.); INRA, UMR 1061 Unité de Génétique Moléculaire Animale, Université de Limoges, IFR 145, Faculté des Sciences et Techniques, 87060 Limoges, France (K.B.).

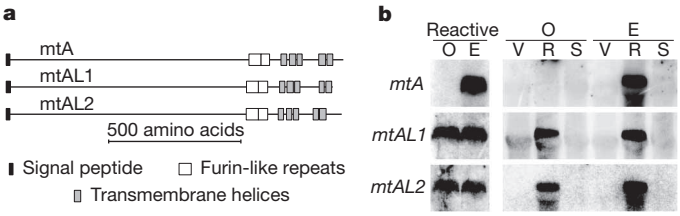


Figure 1 | Structure and expression of mtA and related proteins. **a**, Recognizable protein features. **b**, Northern blot analysis of expression of *mtA*, *mtAL1* (GSPATG00025159001) and *mtAL2* (GSPATG00002922001). The panels on the left compare total RNA samples from sexually reactive cultures of the two mating types (reactive, O and E). In the panels on the right, RNA samples were extracted from: V, vegetative, exponentially growing cells; R, mildly starved and sexually reactive; S, starved for 48 h and no longer reactive. The three mRNAs migrate just above the large rRNA, as can be seen from the pattern of background hybridization.

determined by the rearrangement of one of the genes required for E expression[26].

mtA rearrangements determine mating type

To identify the putative mating-type-determining gene, we used a whole-genome microarray to compare the transcriptomes of sexually reactive cells of both mating types (not shown). The gene found to have the largest E/O expression ratio, GSPATG00017533001, encodes a 1,275-amino-acid protein with a signal peptide and cysteine-rich furin-like repeats followed by 5 transmembrane segments at the carboxy terminus (Fig. 1a and Supplementary Data 1a); the amino-terminal part is predicted to be outside the plasma membrane. This gene was later shown to be *mtA* (ref. 19) (see below). Northern blot analyses confirmed that it is expressed only in E cells and further showed that expression is limited to sexual reactivity, as transcripts could not be detected in exponentially growing cells or in over-starved cells (Fig. 1b). Genes encoding structurally similar proteins were also found to be specifically expressed in sexually reactive cells, but in both mating types (Fig. 1b). RNA interference (RNAi)-mediated silencing of *mtA* in E cells resulted in the default O phenotype during sexual reactivity (as determined by their capacity to agglutinate with E but not with O tester lines), indicating that *mtA* is required for E expression (Supplementary Table 1).

To understand the molecular basis for mating-type-E-specific expression, we sequenced the *mtA* gene from the MACs of O and E cells, and the unrearranged MIC version. The MIC gene is interrupted by four IESs that are excised in both MAC types; however, a 195-base-pair (bp) segment containing the transcription start site and the first 26 bp of the coding sequence was found to be excised as an IES in mating type O MACs, but retained in E MACs (Fig. 2a–c and Supplementary Data 1a). This segment contains the *mtA* promoter, as indicated by microinjection of different constructs into the MACs of O cells (Extended Data Fig. 1a–c): its presence upstream of the coding sequence was sufficient for transformed clones to express mating type E instead of mating type O. An mtA–GFP fusion protein was detected in cilia of the anteroventral surface, although an excess of GFP fluorescence was seen in cytoplasmic structures, probably the ER (Fig. 2d, e). The fusion protein could be detected on cilia with anti-GFP antibodies in fixed, non-permeabilized cells (Fig. 2f), confirming that its N-terminal part is exposed outside of the ciliary membrane.

When *mtA*-transformed clones were taken through autogamy (a self-fertilization sexual process), the induced type E was robustly transmitted to progeny (Extended Data Fig. 1b). As observed for maternally controlled IESs[14,15], a plasmid containing only the 195-bp segment was sufficient to inhibit excision of the homologous sequence during development of zygotic MACs, causing the progeny of transformed O cells to switch to E (Extended Data Fig. 1d). Retention of the *mtA* promoter thus recapitulates both aspects of developmental determination for E, namely the capacity to express mating-type E during the vegetative phase and the capacity, after sexual events, to direct determination of new MACs for the same type.

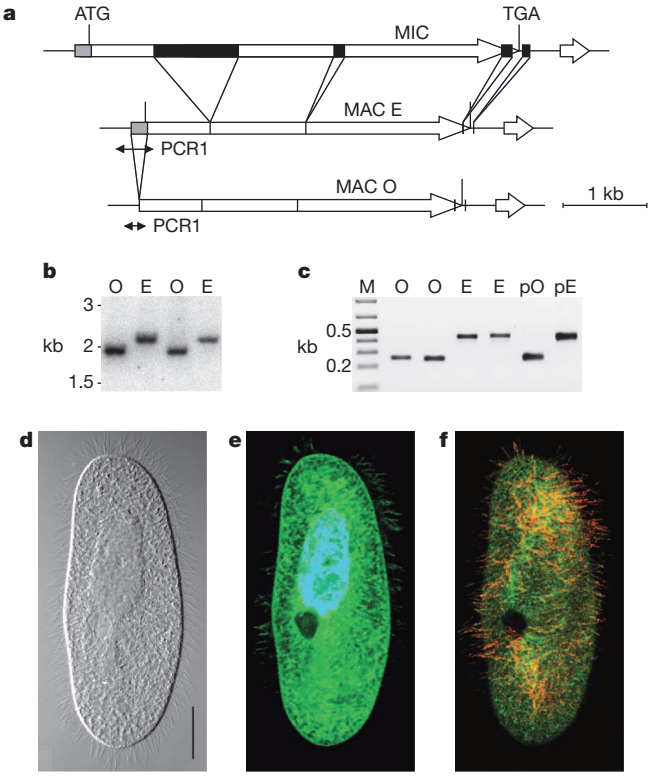


Figure 2 | Structure of *mtA* in the MIC and in the MACs of E and O cells. **a**, Black boxes are IESs; the 195-bp segment (grey box) is excised only in O MACs. A small unannotated gene downstream of *mtA* is conserved in other species. **b**, Southern blot of EcoR1-digested total DNA, hybridized with an *mtA* probe (PCR5) revealing fragments of 1,945 and 2,140 bp in O and E clones. **c**, PCR amplification of the *mtA* 5′ end (PCR1) on independent pools of O and E clones (266 and 461 bp, respectively). pO and pE, control PCRs on cloned MAC versions; M, size markers. **d**–**f**, Localization of an mtA–GFP fusion protein, with GFP inserted between furin-like repeats and transmembrane segments. **d**, DIC image. Scale bar, 20 µm. **e**, Intracellular confocal optical section. Some GFP fluorescence (green) is detected in cilia of the anterior (top), but not posterior, part of the cell. The MAC is stained with Hoechst (blue). **f**, Confocal image of the ventral cell surface, showing the opening of the oral apparatus. The fixed, non-permeabilized cell was labelled with anti-GFP and secondary antibodies (red).

Expression and inheritance uncoupled

Previous mutational analyses showed that the expression of mating types during sexual reactivity can be uncoupled from the mechanism that ensures their transgenerational inheritance[19–21,23]. The recessive mutations mtA^O, mtB^O and mtC^O all preclude expression of type E and restrict cells to the default O phenotype, but they do not affect the maternal inheritance of mating-type determination: mutant homozygotes formed in an E cytoplasmic lineage, despite expressing O, will keep the memory of E determination for an indefinite number of sexual generations, as shown by the fact that their progeny switches back to E expression when the wild-type allele is reintroduced by conjugation[19] (Extended Data Fig. 2a–d).

To determine whether this memory is attributable to retention of the *mtA* promoter, we crossed each of the mutants to the wild type. Autogamy of F_1 heterozygotes of mating-type E yielded homozygous F_2 progeny with a 1:1 ratio of E- and O-expressing clones, reflecting the Mendelian segregation of wild-type and mutant alleles. Both types of clones retained the *mtA* promoter (Extended Data Fig. 2e). That the O-expressing mutant clones could still transmit E determination was verified by crossing them again to wild-type E cells: the second-round F_1 heterozygotes derived from both parents always expressed E and retained the *mtA* promoter (Extended Data Fig. 2f). Autogamy of these F_1 heterozygotes again resulted in the expected Mendelian segregation of E- and O-expressing clones among second-round F_2 homozygotes (Supplementary Table 2). Thus, independently

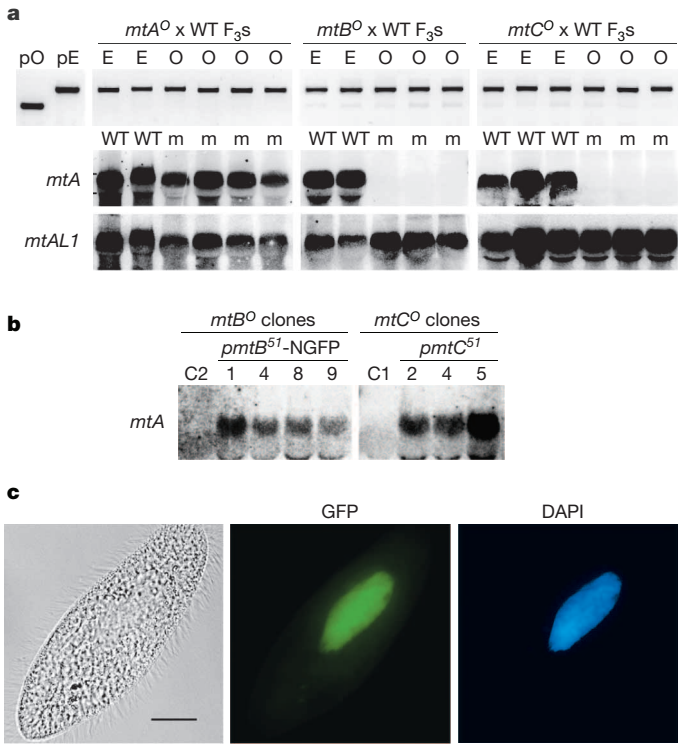


Figure 3 | Molecular analysis of expression mutants. **a**, *mtA* expression in E-determined mtA^O, mtB^O and mtC^O homozygotes. The top panel shows a PCR analysis (PCR1) of E- and O-expressing F_3 lines from the second round of backcross of each mutant to wild-type E cells. pO and pE, control PCRs on cloned MAC versions of *mtA*. All clones retained the *mtA* promoter. The wild-type (WT) or mutant (m) genotypes were confirmed by sequencing. Middle and bottom panels show northern blots of total RNA samples from sexually reactive cultures of each F_3 line, hybridized with *mtA* or *mtAL1* probes. **b**, Northern blot analysis of *mtA* expression in sexually reactive cultures of E-determined mtB^O and mtC^O clones transformed with plasmids $pmtB^{51}$-NGFP and $pmtC^{51}$, respectively (see Extended Data Fig. 3). Transformed clones expressed mating type E, whereas uninjected controls (C2 and C1) expressed mating type O. **c**, Fluorescence microscopy of an mtB^O cell transformed with $pmtB^{51}$-NGFP, expressing mating type E (GFP and DAPI filters). Scale bar, 20 μm.

of the mating type being expressed, maternal inheritance of *mtA* promoter retention may indeed underlie the epigenetic memory of E determination.

To understand how mutant clones can express mating type O despite retaining the *mtA* promoter, we tested *mtA* expression in E- and O-expressing lines, using F_3 populations obtained by an additional autogamy of second-round F_2 homozygotes. In the case of the mtA^O mutation, northern blots showed that *mtA* (GSPATG00017533001) transcripts were produced in sexually reactive cultures of O-expressing lines (Fig. 3a), prompting us to re-sequence the gene in mutant clones. A substitution was found to change Arg codon 751 to a stop codon, and the mutation co-segregated with mating-type O expression among second-round F_2 homozygotes (Supplementary Table 2a). This revealed that GSPATG00017533001 is *mtA* and confirmed that the encoded protein is required for expression of mating type E. The mutation lies at a distance from the *mtA* promoter and apparently does not affect the regulation of its excision, explaining how O-expressing mutant homozygotes can transmit either O or E determination.

In contrast to mtA^O lines, mtB^O and mtC^O O-expressing lines did not produce *mtA* mRNA on sexual reactivity (Fig. 3a). The mutations were identified by whole-genome sequencing, and found to be substitutions in GSPATG00026812001 and GSPATG00009074001, respectively (Supplementary Data 1b, c). Correct identification of the mutations was confirmed by their co-segregation with O expression among F_2 lines (Supplementary Table 2c, d), and by transformation of the MACs of E-determined, O-expressing mutants with the wild-type alleles, which resulted in *mtA* transcription and E expression in both cases (Fig. 3b and Extended Data Fig. 3). Thus, both mtB and mtC seem to be transcription factors required for *mtA* expression in E-determined cells. The *mtB* and *mtC* genes are constitutively expressed at low levels, and both of the encoded proteins are predicted to be nuclear. mtC is a 138-amino-acid protein containing a C2H2 zinc finger, the structure of which is probably affected by the mtC^O mutation (Supplementary Data 2). Although the 310-amino-acid mtB protein does not contain any recognizable domain, an mtB–GFP fusion protein was found to localize to the MAC (Fig. 3c). Deep sequencing of the transcriptomes of mtB^O and mtC^O mutants further showed that *mtA* is the only gene that requires both factors for expression during sexual reactivity; importantly, it is the only gene found to be downregulated in the mtB^O mutant (Supplementary Table 3).

mtA promoter excision is regulated by scnRNAs

Excision of the *mtA* promoter in mating-type O cells occurs between two 5′-TA-3′ dinucleotides within a reasonable IES end consensus (Supplementary Data 1a), suggesting that the general IES excision machinery is involved. RNAi-mediated depletion of the Pgm endonuclease[5] during autogamy of O cells indeed impaired excision and resulted in accumulation of unexcised copies in the developing new MAC (Fig. 4a). So did depletion of proteins known to be involved in scnRNA biogenesis or action, including the Dicer-like Dcl2 and Dcl3 (ref. 12), the Piwi-like Ptiwi01 and Ptiwi09 (ref. 11), and the Nowa1 and Nowa2 RNA-binding proteins[27], indicating that the scnRNA pathway is required to target excision in O cells. Depletion of each pair of proteins causes massive retention of IESs genome-wide, resulting in non-functional new MACs and post-autogamous lethality. In conditions of partial depletion, obtained by silencing only one of the two Piwi (or the two Dicer-like) genes, the progeny is usually viable. Analysis of individual clones showed that they frequently retained the *mtA* promoter in the new MAC, and in all cases this correlated with a switch to mating type E (Fig. 4b).

No other IES is known to be affected in these conditions; in this regard, excision of the *mtA* promoter behaves like MAC deletions of cellular genes, which were shown to be more sensitive to partial impairment of the scnRNA pathway[11,12]. Whereas other IESs are intervening sequences

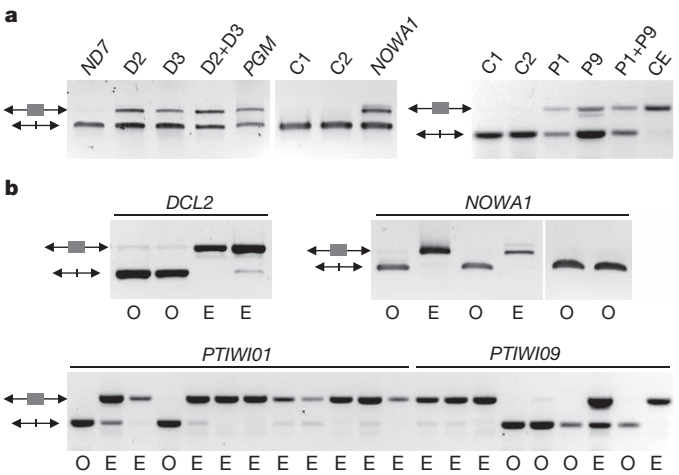


Figure 4 | Genes required for excision of the *mtA* promoter in O cells. **a**, PCR analysis (PCR5) of *mtA* promoter retention in mass progenies of O clones after RNAi-mediated silencing of the indicated genes. Total DNA samples were prepared from starved post-autogamous cells when the new MACs were clearly visible. Because the parental MAC is still present at this stage, the promoter-excised version is amplified in all cases; the promoter-retaining fragment can be detected only if it accumulates in zygotic MACs. *ND7*, unrelated-gene negative control; D2, *DCL2*; D3, *DCL3*; C1, no-silencing negative control; C2, empty-vector RNAi control; P1, *PTIWI01*; P9, *PTIWI09*; CE, control PCR5 on E cells. **b**, PCR analysis (PCR5) and mating-type tests (below) of individual viable post-autogamous clones from non-lethal silencing conditions (including partial silencing of *NOWA1*).

that must be excised to reconstitute functional genes, the 195-bp segment is a functional part of the *mtA* gene. The homologous segment is not excised in *mtA* orthologues from other *P. aurelia* species (see below), indicating that excision of a bona fide part of the MAC genome evolved as a derived character in *P. tetraurelia*, as random mutations happened to create sites suitable for Pgm-mediated excision.

Deep sequencing of small RNAs from an early conjugation time point confirmed that scnRNAs are produced from both strands of the *mtA* promoter, as they are from the rest of the germline genome (Extended Data Fig. 4). On average, scnRNA coverage was ~2-fold higher for IESs than for other genome features, as independently found in another study[28]. Although we cannot exclude that this results from precocious degradation of MAC-matching scnRNAs, an intrinsically more abundant production of scnRNAs from IESs could explain why IES excision is less sensitive to partial impairment of the scnRNA pathway than gene deletion. Northern blot analyses of autogamy time courses showed that *mtA*-promoter scnRNAs are produced in similar amounts during meiosis of O and E cells, and failed to detect any difference in the timing of their disappearance (Extended Data Fig. 5). Thus, *mtA*-promoter scnRNAs may be inactivated in E cells by some other mechanism than the active degradation proposed in *Tetrahymena thermophila*[29–31], for instance by sequestration in the maternal MAC. Whatever the precise mechanism, one prediction of the model is that RNAi-mediated destruction of maternal transcripts of the *mtA* promoter during autogamy of E cells will prevent inactivation of homologous scnRNAs, licensing them to target excision in the zygotic MAC[13]. Indeed, feeding E cells with double-stranded RNA homologous to the *mtA* promoter before autogamy induced its precise excision in the new MAC and resulted in O progeny (Supplementary Table 4). The induced type O was thereafter inherited for at least three sexual generations (not shown).

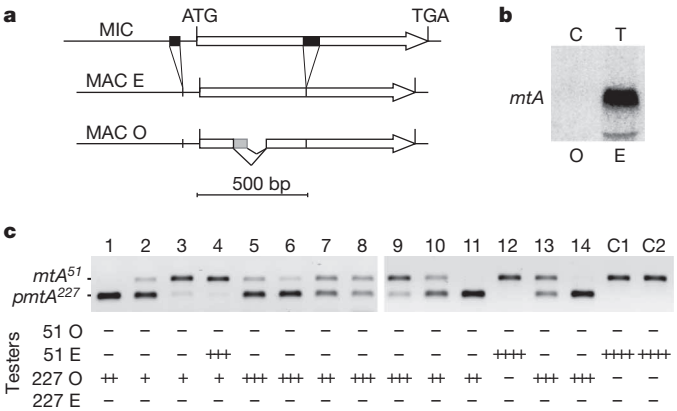


Figure 5 | *mtB* rearrangements determine mating types in *P. septaurelia*. **a**, MIC and MAC versions of mtB^{227}. The MIC sequence contains two IESs (black boxes), conserved in *P. tetraurelia* strain 51. In the MAC of O cells of strain 227, the mtB^{227} coding sequence suffers two alternative deletions, each between TA dinucleotides within IES-end consensus sequences (TAYAG, see Supplementary Data 1b). **b**, Northern blot analysis of *mtA* expression in a sexually reactive, $pmtB^{51}$-transformed O clone of strain 227 (T). C, control uninjected O clone. Tested mating types are indicated. **c**, Transformation of *P. tetraurelia* clones of mating type O with the *P. septaurelia* mtA^{227} gene. A duplex PCR was used to assess transgene copy numbers through the relative abundance of a 176-bp transgene-specific product ($pmtA^{227}$, PCR12), compared to the 266-bp product from the promoterless endogenous *mtA* (mtA^{51}, PCR1; undetectable in high-copy transformants). C1 and C2, uninjected control clones. Each clone was tested with *P. tetraurelia* (51) and *P. septaurelia* (227) O and E testers. Number of plus signs represents strength of mating reaction; minus indicates no agglutination.

A different switch in *P. septaurelia*

P. tetraurelia belongs to a group of 15 sibling species that are morphologically indistinguishable but sexually incompatible[1,32]. All have homologous O and E mating types, as shown by cross-agglutination between the O types of some species and the E types of others. However, maternal inheritance of mating types is observed in only half of these species. In others, mating types are randomly determined during MAC development, without any influence of the maternal MAC. The distribution of these systems in the phylogenetic tree of *aurelia* species suggests multiple changes during evolution of the complex[33,34] (Extended Data Fig. 6). We first sequenced *mtA*, *mtB* and *mtC* orthologues in two strains of *P. octaurelia*, a maternal-inheritance species closely related to *P. tetraurelia*, with which it can form viable (though sterile) F_1 hybrids[35]. We found the *mtA* promoter to be excised in O clones, with the same boundaries as in *P. tetraurelia*, but not in E clones (Extended Data Fig. 7 and Supplementary Data 1a). This excision system thus probably evolved in the common ancestor of the two species.

We then examined *P. septaurelia*, another maternal-inheritance species that groups with random-determination species. Sequencing of the *mtA*, *mtB* and *mtC* orthologues in the wild-type strain 227 showed that the mtA^{227} promoter was not excised in the MACs of O cells (Supplementary Data 1a). Instead, MAC copies of the mtB^{227} gene contained either of two alternative deletions of coding-sequence segments between IES-like boundaries (Fig. 5a and Supplementary Data 1b), suggesting that the gene was rearranged into non-functional forms. Our stock of strain 227 contained only O cells, but complementation of their MACs by microinjection of the *P. tetraurelia* $pmtB^{51}$ transgene made these cells phenotypically E and resulted in *mtA* transcription during sexual reactivity (Fig. 5b). Thus, mating type E is characterized by mtB-dependent *mtA* expression in *P. septaurelia*, as in *P. tetraurelia*.

The *mtB* gene belongs to a family of up to six paralogues in *P. aurelia* species, and sequence conservation indicates that all have been under purifying selection since before speciation (not shown). The excision events that inactivate mtB^{227} in O cells are not observed in any other paralogue or in *mtB* orthologues from other species, where the 5′ TAs correspond to different nucleotides. As with *mtA* promoter excision in *P. tetraurelia*, this excision system apparently does not originate from any transposable element insertion but seems to have arisen in *P. septaurelia* after the chance appearance of good matches to the IES end consensus, which seems to be conserved among *P. aurelia* species[36].

Taking advantage of the evidence that the mtB^{51} protein from *P. tetraurelia* can activate the mtA^{227} promoter in *P. septaurelia*, we microinjected the mtA^{227} gene into the MACs of *P. tetraurelia* cells of mating type O, and tested the sexual preference of transformed clones with O and E testers of both species (Fig. 5c). Consistent with the reported lack of cross-agglutination between the complementary types of these two species[1], uninjected control clones reacted only with *P. tetraurelia* E testers. In contrast, transformed clones agglutinated only with *P. septaurelia* O testers, allowing interspecific conjugation to occur. Thus, mtA^{227} is sufficient for expression of the *P. septaurelia* E specificity, incidentally confirming that mtA localizes on the external side of ciliary membranes. The lack of reaction of transformed clones with *P. tetraurelia* O testers indicates that mtA interacts with an unknown O-specific receptor which also differs between species. Furthermore, transformed clones did not react with *P. tetraurelia* E testers, indicating that mtA^{227} expression, like mtA^{51} expression (Extended Data Fig. 1), functionally masks expression of the endogenous O-specific receptor. This would explain the default nature of the O type.

Strain 38 is a natural *P. septaurelia* isolate carrying a Mendelian mutation at a single locus (*mt*) which, in contrast to the *P. tetraurelia* mutations, affects both the expression and the epigenetic inheritance of mating types[20]. This dual effect suggests that *mt* is the gene that controls mating types through an alternative rearrangement in that species (see Extended Data Fig. 8), providing an independent means to confirm the identification of the *P. septaurelia* switch. Sequencing of the mtB^{38} allele revealed mutations that can explain both effects (Extended Data Fig. 8c and Supplementary Data 1b), and the analysis of a cross between strains 38 and 227 showed that *mt* is identical or closely linked to *mtB* (Supplementary Table 5). To confirm directly that *mtB* rearrangements are regulated by homology-dependent maternal effects, the full-length, functional MAC form

of the *mtB*^{227} allele was microinjected into the MACs of O cells of strain 227. Transformed clones expressed E and produced E post-autogamous progeny (Extended Data Fig. 9). We conclude that mating types are determined in *P. septaurelia* by maternally inherited alternative rearrangements of the *mtB* gene, a mechanism that must have evolved independently from that of *P. tetraurelia*.

Discussion

Seventy-five years after mating types were first described in *P. tetraurelia*, the expression of a single gene, *mtA*, was found to make the difference between the two types. *mtA* transcription in E cells was further shown to require the *mtB* and *mtC* gene products, and this E-specific pathway seems to be conserved in other *P. aurelia* species. Our results indicate that mtA is a ciliary transmembrane protein directly involved in the species-specific recognition of O cells. Although the nature of the O-specific receptor remains unknown, its expression or function appears to be masked by the mtA protein, given the default O phenotype observed in mutants. Expressing the *P. septaurelia* mtA orthologue in O cells of *P. tetraurelia* not only provided them with the *P. septaurelia* E specificity, but also blocked expression of the *P. tetraurelia* O specificity. In the related *P. caudatum*, an antigen involved in agglutination was shown to be produced in the cytoplasm of both O and E reactive cells, but to localize in the ciliary membrane only in O cells[37]. In E cells of *P. aurelia* species, the mtA protein could inhibit the synthesis, processing or transport of the O-specific receptor. mtA is structurally similar to the proteins recently shown to be specific for each of the seven mating types of *T. thermophila*[38], but how these function in mating-type recognition remains to be determined.

Despite conservation of the *mtA* on/off switch as the E/O expression determinant, the mechanisms used to produce O clones differ among *P. aurelia* species. Excision of the *mtA* promoter is shared between the closely related *P. tetraurelia* and *P. octaurelia*, but in *P. septaurelia mtA* expression is prevented by deletions in *mtB* (Extended Data Fig. 10). This seems to be functionally equivalent, as mtB is highly specific for the *mtA* promoter in *P. tetraurelia*. The *P. tetraurelia* rearrangement involves the IES excision machinery and is regulated by the scnRNA pathway, explaining the maternal inheritance of mating types in this species, and probably in the other two as well. Ciliate scnRNAs resemble metazoan piRNAs in many respects, including their central roles in epigenetic reprogramming events that are critical for sexual reproduction[39], and both have been described as genomic 'immune systems' primarily involved in the control of transposable elements[2,40–42]. One difference, however, is that metazoan piRNA systems keep a memory of 'non-self' sequences, based on integration into specific small-RNA-producing loci[40,42,43]. The *Paramecium* system seems to work the opposite way, keeping a memory of 'self' sequences in the form of the maternal MAC, which is probed by genome-wide scnRNAs at each generation to identify 'non-self' molecules by subtraction.

As a consequence, the *Paramecium* genome-defence mechanism can easily be recruited for heritable regulation of bona fide cellular genes, an important question that has been raised in diverse eukaryotes[42,44–46]. Indeed, the sequences being excised from the *mtA* and *mtB* genes differ from other IESs[3,7] in that they are functional parts of cellular genes and apparently do not derive from the insertion of transposable elements; moreover, the *mtA* rearrangement behaves like deletions of cellular genes in its mechanistic requirements, which show subtle differences with IES excision[11,12]. The independent evolution of the *P. tetraurelia* and *P. septaurelia* switches suggests that exaptation of the scnRNA pathway is a general mechanism for transgenerational epigenetic inheritance of differentiated states. In plants and mammals, epigenetic regulation of cellular genes often relies on transposon-induced epigenetic changes[44,45], as is the case of sex determination in melon[47]. The capacity of the *Paramecium* mechanism to inactivate any single-copy gene makes it even more flexible, dispensing with the need for transposon insertion. In keeping with the low information content of the IES end consensus, accidental rearrangement errors are not rare genome-wide[3,36,48], and may occasionally provide a selective advantage in specific conditions; maternal transmission to sexual progeny would thus allow continuous adaptation of the somatic (MAC) genome, independently of any Mendelian (MIC) mutation.

METHODS SUMMARY

Unless otherwise stated, all experiments were carried out with the entirely homozygous strain 51 of *P. tetraurelia*; the origins of other strains and species are given in the Methods section. Molecular biology experiments used standard procedures. *Paramecium*-specific methods (RNAi-mediated gene silencing by dsRNA feeding, DNA microinjection into the MAC) have been published and are referenced in the Methods section. For mating-type tests, testers were prepared from cell lines of known mating types by re-feeding ~1,000 autogamous cells in tubes for 2–3 divisions. The procedures used to test mass post-autogamous progenies or individual clones are presented in the Methods section. Bioinformatic analyses and statistical testing (microarray expression data, whole-genome transcriptome profiling by RNA-seq) are described in the Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 August 2013; accepted 11 April 2014.
Published online 7 May 2014.

1. Sonneborn, T. M. Paramecium aurelia. in *Handbook of Genetics* (ed. King, R. C.) 469–594 (Plenum, 1974).
2. Chalker, D. L. & Yao, M. C. DNA elimination in ciliates: transposon domestication and genome surveillance. *Annu. Rev. Genet.* **45,** 227–246 (2011).
3. Arnaiz, O. *et al.* The *Paramecium* germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS Genet.* **8,** e1002984 (2012).
4. Bétermier, M. Large-scale genome remodelling by the developmentally programmed elimination of germ line sequences in the ciliate *Paramecium*. *Res. Microbiol.* **155,** 399–408 (2004).
5. Baudry, C. *et al.* PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes Dev.* **23,** 2478–2483 (2009).
6. Klobutcher, L. A. & Herrick, G. Consensus inverted terminal repeat sequence of *Paramecium* IESs: resemblance to termini of Tc1-related and Euplotes Tec transposons. *Nucleic Acids Res.* **23,** 2006–2013 (1995).
7. Dubois, E. *et al.* Transposon invasion of the *Paramecium* germline genome countered by a domesticated PiggyBac transposase and the NHEJ pathway. *Int. J. Evol. Biol.* **2012,** 436196 (2012).
8. Chalker, D. L., Meyer, E. & Mochizuki, K. Epigenetics of ciliates. *Cold Spring Harb. Perspect. Biol.* **5,** a017764 (2013).
9. Coyne, R. S., Lhuillier-Akakpo, M. & Duharcourt, S. RNA-guided DNA rearrangements in ciliates: is the best genome defence a good offence? *Biol. Cell* **104,** 309–325 (2012).
10. Duharcourt, S., Lepere, G. & Meyer, E. Developmental genome rearrangements in ciliates: a natural genomic subtraction mediated by non-coding transcripts. *Trends Genet.* **25,** 344–350 (2009).
11. Bouhouche, K., Gout, J. F., Kapusta, A., Betermier, M. & Meyer, E. Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Res.* **39,** 4249–4264 (2011).
12. Lepere, G. *et al.* Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Res.* **37,** 903–915 (2009).
13. Lepere, G., Betermier, M., Meyer, E. & Duharcourt, S. Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in *Paramecium tetraurelia*. *Genes Dev.* **22,** 1501–1512 (2008).
14. Duharcourt, S., Butler, A. & Meyer, E. Epigenetic self-regulation of developmental excision of an internal eliminated sequence on *Paramecium tetraurelia*. *Genes Dev.* **9,** 2065–2077 (1995).
15. Duharcourt, S., Keller, A. M. & Meyer, E. Homology-dependent maternal inhibition of developmental excision of internal eliminated sequences in *Paramecium tetraurelia*. *Mol. Cell. Biol.* **18,** 7075–7085 (1998).
16. Garnier, O., Serrano, V., Duharcourt, S. & Meyer, E. RNA-mediated programming of developmental genome rearrangements in *Paramecium tetraurelia*. *Mol. Cell. Biol.* **24,** 7370–7379 (2004).
17. Sonneborn, T. M. Sex, sex inheritance and sex determination in *Paramecium aurelia*. *Proc. Natl Acad. Sci. USA* **23,** 378–385 (1937).
18. Butzel, H. M. Mating type mutations in variety 1 of *Paramecium aurelia*, and their bearing upon the problem of mating type determination. *Genetics* **40,** 321–330 (1955).
19. Byrne, B. C. Mutational analysis of mating type inheritance in Syngen 4 of *Paramecium aurelia*. *Genetics* **74,** 63–80 (1973).
20. Taub, S. R. The genetic control of mating type differentiation in *Paramecium*. *Genetics* **48,** 815–834 (1963).
21. Sonneborn, T. M. Genetics of cellular differentiation: stable nuclear differentiation in eucaryotic unicells. *Annu. Rev. Genet.* **11,** 349–367 (1977).
22. Sonneborn, T. M. Recent advances in the genetics of *Paramecium* and Euplotes. *Adv. Genet.* **1,** 263–358 (1947).

23. Nanney, D. L. Mating type inheritance at conjugation in variety 4 of *Paramecium aurelia*. *J. Protozool.* **4,** 89–95 (1957).
24. Sonneborn, T. M. Patterns of nucleo-cytoplasmic integration in *Paramecium*. *Caryologia* **6** (suppl.), 307–325 (1954).
25. Meyer, E. & Keller, A. M. A Mendelian mutation affecting mating-type determination also affects developmental genomic rearrangements in *Paramecium tetraurelia*. *Genetics* **143,** 191–202 (1996).
26. Meyer, E. & Garnier, O. Non-mendelian inheritance and homology-dependent effects in ciliates. *Adv. Genet.* **46,** 305–337 (2002).
27. Nowacki, M., Zagorski-Ostoja, W. & Meyer, E. Nowa1p and Nowa2p: novel putative RNA binding proteins involved in trans-nuclear crosstalk in *Paramecium tetraurelia*. *Curr. Biol.* **15,** 1616–1628 (2005).
28. Sandoval, P. Y., Swart, E. C., Arambasic, M. & Nowacki, M. Functional diversification of Dicer-like proteins and small RNAs required for genome sculpting. *Dev. Cell* **28,** 174–188 (2014).
29. Aronica, L. *et al.* Study of an RNA helicase implicates small RNA-noncoding RNA interactions in programmed DNA elimination in *Tetrahymena*. *Genes Dev.* **22,** 2228–2241 (2008).
30. Mochizuki, K. & Gorovsky, M. A. Conjugation-specific small RNAs in *Tetrahymena* have predicted properties of scan (scn) RNAs involved in genome rearrangement. *Genes Dev.* **18,** 2068–2073 (2004).
31. Schoeberl, U. E., Kurth, H. M., Noto, T. & Mochizuki, K. Biased transcription and selective degradation of small RNAs shape the pattern of DNA elimination in *Tetrahymena*. *Genes Dev.* **26,** 1729–1742 (2012).
32. Catania, F., Wurmser, F., Potekhin, A. A., Przybos, E. & Lynch, M. Genetic diversity in the *Paramecium aurelia* species complex. *Mol. Biol. Evol.* **26,** 421–431 (2009).
33. Hall, M. S. & Katz, L. A. On the nature of species: insights from *Paramecium* and other ciliates. *Genetica* **139,** 677–684 (2011).
34. Phadke, S. S. & Zufall, R. A. Rapid diversification of mating systems in ciliates. *Biol. J. Linn. Soc.* **98,** 187–197 (2009).
35. Haggard, B. W. Interspecies crosses in *Paramecium aurelia* (syngen 4 by syngen 8). *J. Protozool.* **21,** 152–159 (1974).
36. Catania, F., McGrath, C. L., Doak, T. G. & Lynch, M. Spliced DNA sequences in the *Paramecium* germline: their properties and evolutionary potential. *Genome Biol. Evol.* **5,** 1200–1211 (2013).
37. Xu, X., Kumakura, M., Kaku, E. & Takahashi, M. Odd mating-type substances may work as precursor molecules of even mating-type substances in *Paramecium caudatum*. *J. Eukaryot. Microbiol.* **48,** 683–689 (2001).
38. Cervantes, M. D. *et al.* Selecting one of several mating types through gene segment joining and deletion in *Tetrahymena thermophila*. *PLoS Biol.* **11,** e1001518 (2013).
39. Bourc'his, D. & Voinnet, O. A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. *Science* **330,** 617–622 (2010).
40. Malone, C. D. & Hannon, G. J. Small RNAs as guardians of the genome. *Cell* **136,** 656–668 (2009).
41. Schoeberl, U. E. & Mochizuki, K. Keeping the soma free of transposons: programmed DNA elimination in ciliates. *J. Biol. Chem.* **286,** 37045–37052 (2011).
42. Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. A. PIWI-interacting small RNAs: the vanguard of genome defence. *Nature Rev. Mol. Cell Biol.* **12,** 246–258 (2011).
43. Khurana, J. S. *et al.* Adaptation to P element transposon invasion in *Drosophila melanogaster*. *Cell* **147,** 1551–1563 (2011).
44. Castel, S. E. & Martienssen, R. A. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nature Rev. Genet.* **14,** 100–112 (2013).
45. Daxinger, L. & Whitelaw, E. Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nature Rev. Genet.* **13,** 153–162 (2012).
46. Luteijn, M. J. & Ketting, R. F. PIWI-interacting RNAs: from generation to transgenerational epigenetics. *Nature Rev. Genet.* **14,** 523–534 (2013).
47. Martin, A. *et al.* A transposon-induced epigenetic change leads to sex determination in melon. *Nature* **461,** 1135–1138 (2009).
48. Duret, L. *et al.* Analysis of sequence variability in the macronuclear DNA of *Paramecium tetraurelia*: a somatic view of the germline. *Genome Res.* **18,** 585–596 (2008).
49. Edgar, R., Domrachev, M. & Lash, A. E. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30,** 207–210 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Malinsky, C. Ciaudo and M.-A. Félix for critical reading of the manuscript, and S. Marker and all other laboratory members for continuous support and discussions. This work was supported by the 'Investissements d'Avenir' program ANR-10-LABX-54 MEMO LIFE/ANR-11-IDEX-0001-02 Paris Sciences et Lettres* Research University and by grants ANR-08-BLAN-0233 'ParaDice' and ANR-12-BSV6-0017 'INFERNO' to E.M., L.S. and S.D., an 'Equipe FRM' grant to E.M., grants ANR-2010-BLAN-1603 'GENOMAC' and CNRS ATIP-Avenir to S.D., and National Science Foundation grant MCB-1050161 to M. Lynch (JFG). D.P.S. was supported by Ph.D. fellowships from the Erasmus Mundus program and from the Ligue Nationale Contre le Cancer. M.L.-A. was supported by Ph.D. fellowships from the Ministère de l'Enseignement Supérieur et de la Recherche and from the Fondation de la Recherche Médicale. A.P. was supported by grant RFBR 13-04-01683a. Some strains used in this study are maintained at the Centre of Core Facilities 'Culture Collection of Microorganisms' in St Petersburg State University. The sequencing of the mtB^O and mtC^O MAC genomes benefited from the facilities and expertise of the high-throughput sequencing platform of IMAGIF (Centre de Recherche de Gif, http://www.imagif.cnrs.fr). The mtB^O and mtC^O transcriptomes were sequenced at the Genomic Paris Centre - IBENS platform, member of 'France Gènomique' (ANR10-INBS-09-08). This study was carried out in the context of the CNRS-supported European Research Group 'Paramecium Genome Dynamics and Evolution' and the European COST Action BM1102.

Author Contributions D.P.S. did almost all of the experimental work presented here and contributed to the design of experiments. B.S. characterized mRNAs and contributed to silencing experiments and northern blot analyses. G.G. contributed to gene sequencing, plasmid construction, PCR analyses and cell line maintenance. J.-F.G. did the microarray analysis, and A.A.-F. the confocal analysis of mtA–GFP fusions. A.A., K.L. and J.-M.A. carried out the deep sequencing of small RNAs, and C.B. that of the mtB^O and mtC^O transcriptomes; O.A. and L.S. did the bioinformatic analyses. K.B., M.L.-A., V.T. and S.D. showed the role of scnRNA pathway genes in *mtA* promoter excision. S.B. did the *mtA* promoter dsRNA feeding experiment. A.P. contributed to the analysis of the mtA^O mutant and provided *P. octaurelia* and *septaurelia* strains. M.P. contributed to the analysis of the mtB^O mutant and prepared samples from the *P. octaurelia* cross, which was carried out by E.P. E.M. conceived the study and wrote the paper.

Author Information Microarray data have been deposited at the Gene Expression Omnibus database[49] under accession number GSE43436. RNA-seq data (transcriptomes of mtB^O and mtC^O mutants) have been deposited in the European Nucleotide Archive (EBI) under accession number ERP002291. Small RNA sequences have been deposited at the EBI under accession number ERP001812. The *mtA*, *mtB* and *mtC* sequences of all strains and species studied have been deposited at GenBank under accession codes KJ748544–KJ748569. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.M. (emeyer@biologie.ens.fr).

METHODS

***Paramecium* strains and cultivation.** Unless otherwise stated, all experiments were carried out with the entirely homozygous strain 51 of *P. tetraurelia*. The mutant strains *mtA*^{O}, *mtB*^{O} and *mtC*^{O}, as well as strain 32 of *P. tetraurelia*, were from the stock collection of the Centre de Génétique Moléculaire in Gif-sur-Yvette, France. The *mtA*^{O} mutant was independently ordered from ATCC (ATCC number 30762). Strains 138 and GFg-1 of *P. octaurelia* were from the stock collection of the Institute of Systematics and Evolution of Animals, Krakow, Poland. Strain 227 from *P. septaurelia* was from the stock collection of the Centre of Core Facilities 'Cultivation of Microorganisms' in St Petersburg State University, Russia. Strain 38 of *P. septaurelia* was ordered from ATCC (ATCC number 30575). Cultivation and autogamy were carried out at 27 °C (unless otherwise stated) as described[50,51].

DNA and RNA extraction, Southern and northern blots. DNA and RNA samples were typically extracted from 200- to 400-ml cultures of exponentially growing cells at <1,000 cells ml^{-1} or of autogamous cells at 2,000–4,000 cells ml^{-1} as previously described[5]. Small-scale DNA samples were prepared from ≤1,000 cells using the NucleoSpin Tissue kit (Macherey-Nagel). The TRIzol (Invitrogen) RNA extraction procedure was adapted for small-scale cultures (~20 ml) of individual transformed clones during sexual reactivity. Electrophoresis and blotting were carried out according to standard procedures. For small-RNA northern blots, small RNAs were enriched by PEG8000 precipitation, and the equivalent of 75 µg of total RNA for each time point was run on a 15% polyacrylamide-urea gel, transferred on Hybond NX membranes, and chemically crosslinked[52].

Microarray expression data. Expression data were obtained from single-channel NimbleGen microarrays covering all 39,642 annotated genes, with six different 50-mer probes per gene. Raw signals were processed using the standard RMA method[53]. This includes a first step of background subtraction for each array, followed by between-array normalization which was carried out using the normalizeBetweenArrays function from the limma package[54]. The latter step adjusts signals so that expression values have similar distributions in the two arrays considered in the analysis. The expression level of each gene was taken as the median signal from the six probes. The microarray platform has been described in more detail elsewhere[55].

Alignment and prediction programs. DNA and protein sequences were aligned using the MUSCLE software on the phylogeny.fr website[56]. Prediction of protein localization and transmembrane protein topology used the PSORT II[57] and PolyPhobius[58] servers.

Mating-type tests. Testers were prepared from cell lines of known mating types by re-feeding ~1,000 autogamous cells in tubes with 4 ml of 0.2 X WGP medium bacterized with *Klebsiella pneumoniae* (light medium) and incubating overnight at 27 °C. The next day, tubes were re-fed with 8 ml of light medium and again incubated overnight at 27 °C. The following day, reactive cells concentrated near the top of the tube were collected (~1.5 ml per tube), checked by mixing aliquots with reactive cells of the complementary type, and used in mating type tests. Mass post-autogamous progenies to be tested were made reactive in the same way. To test individual clones, single karyonides (out of autogamy or conjugation) were isolated in 250 µl of light medium and incubated until starved. They were then re-fed with 250 µl and tested the next day.

RNAi-mediated gene silencing by dsRNA feeding and microinjections. Procedures for RNAi and DNA microinjection into the MAC, as well as inserts used to silence scnRNA-pathway genes and *PGM* and the procedure to silence genes during autogamy, have been described[5,11,12,27,59,60]. To silence genes during sexual reactivity, cells were made reactive as described under 'Mating-type tests', except that the second re-feeding with 8 ml was done with dsRNA feeding medium instead of normal WGP medium.

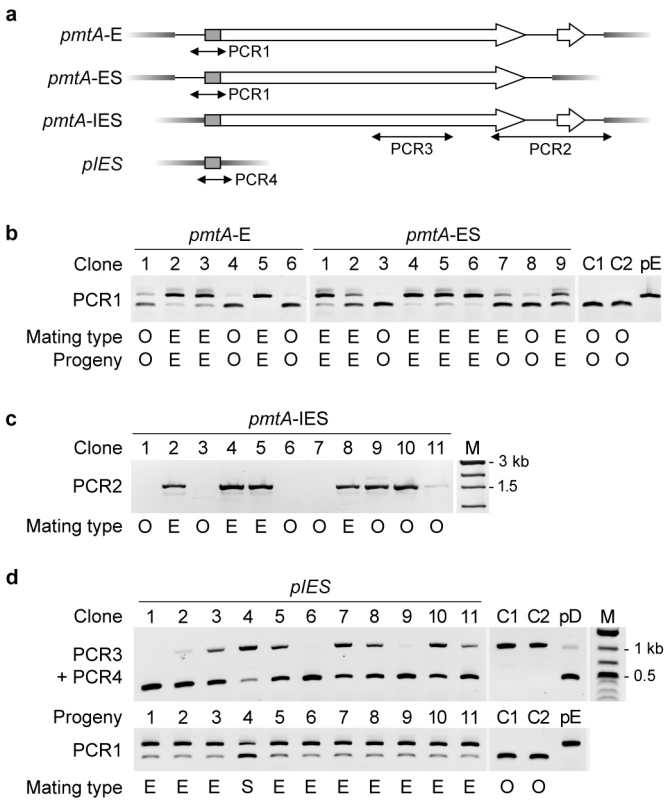
Plasmid constructs and probes. The plasmids used in this study are described in the relevant figures or Extended Data. The *mtA*-GFP fusion transgene used in Fig. 2d–f is a modified version of pmtA-ES (Extended Data Fig. 1) in which the EGFP coding sequence, flanked by Gly-Ser-Gly-Gly and Gly-Gly linkers, was inserted in place of mtA residue Asn 1002. The *pmtA*^{227} transgene used in Fig. 5c is the *P. septaurelia* equivalent of pmtA-ES, with complete flanking intergenic regions (without the downstream unannotated gene). Complete sequences of all plasmids are available on request. The *mtA* probe used to hybridize northern blots covered the whole coding sequence, except on the left panel of Fig. 1c where it spanned positions 415–1010 of the corrected gene model (PTETG5300016001, see ParameciumDB). The *P. septaurelia* probe in Fig. 5b and Extended Data Fig. 8d covered region 78–2245 of the *mtA*^{38} gene model. For other genes, probes covered the indicated regions of each gene model: *mtAL1*, 3053–3599 of GSPATG00025159001; *mtAL2*, 3328–3790 of GSPATG00002922001; *PTIWI09*, 50–439 of GSPATG00020796001; *NOWA2*, 2513–3330 of GSPATG00016668001; *PGM*, 1540–2390 of GSPATG00016627001; *PTIWI10*, 4–264 of GSPATG00009468001. Oligonucleotide probes for scaffold-22 endogenous siRNAs (Extended Data Fig. 5c) have been described[61].

Confocal analysis of mtA–GFP localization. Cells were fixed in 2% paraformaldehyde in PBS buffer and rinsed in the same buffer supplemented with 3% BSA. All subsequent steps were performed in this buffer: cells were incubated with a commercial anti-GFP rabbit serum (Molecular Probes A6455, Invitrogen) at a 1/1,000 dilution, rinsed twice, incubated with an Alexa 568-coupled anti-rabbit antibody (Molecular Probes goat anti-rabbit IgG(H+L) A11011, Invitrogen), rinsed in 1 µg ml^{-1} Hoechst 33258, mounted in Citifluor (Citifluor Ltd London), and observed under a Leica Confocal SP8 microscope with 405, 488 and 552 laser line excitations for blue, GFP and DIC, and red detection, respectively. Images were merged using ImageJ and Adobe Photoshop.

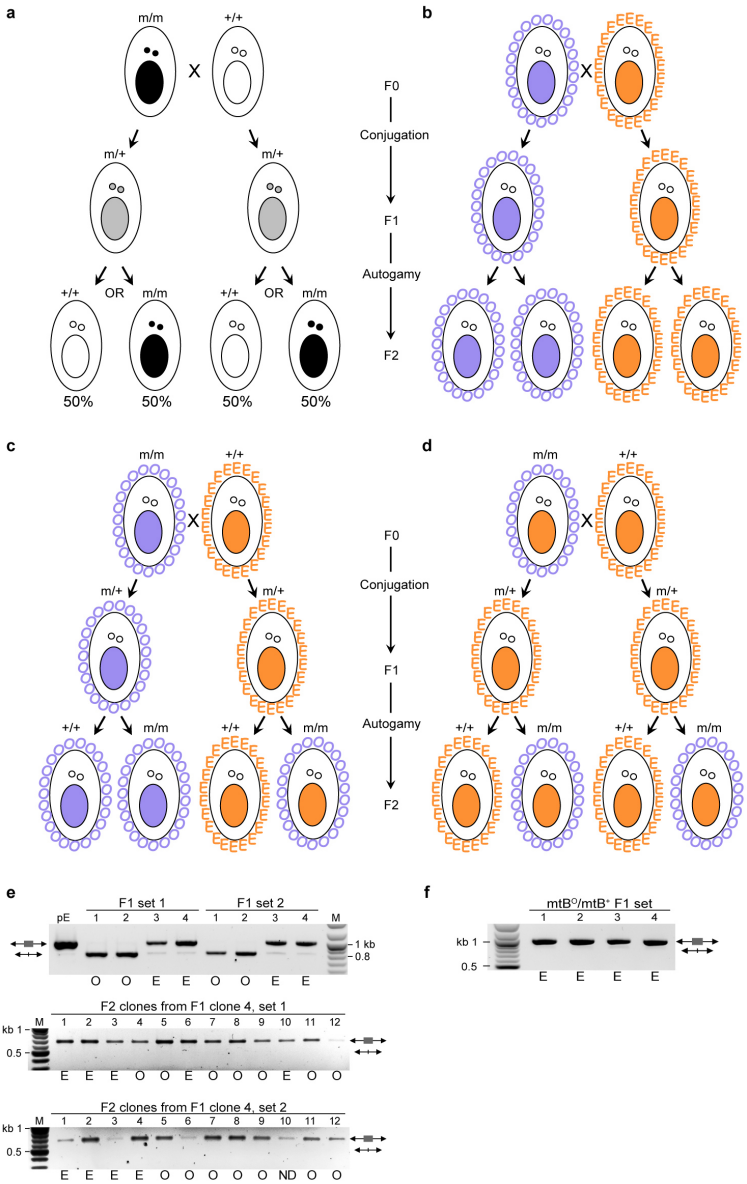
Deep sequencing of scnRNAs. Small-RNA libraries were previously constructed using 24-nt adaptors[12]. The corresponding PCR products were extended with 6 degenerate nucleotides at each end and ligated to Illumina adaptors essentially following Illumina's recommendations (TruSeq DNA sample prep kit protocol) before sequencing on the GAIIx system (100-nt single reads) (Illumina, USA).

Whole-genome transcriptome profiling. Eight Illumina sequencing libraries were constructed (two biological replicates for *mtB*^{O}, two for *mtC*^{O}, and two wild-type controls for each) using the Illumina TruSeq RNA Sample Preparation kit. Approximately 30 million 101-nt read pairs were obtained for each library with the Illumina HiSeq 1000 sequencing system. Reads were processed to remove adapters and trimmed for quality before mapping to the strain 51 genome[3,62] using BWA software[63]. The fragments that mapped to each annotated gene[62] were counted only if both reads in the pair mapped to the same gene and only if the mapping for both reads was unique in the genome. Fragments were only counted if neither read in the pair had more than two mismatches. Fragment counts per gene were obtained from the BWA output using Samtools[64] and a custom Perl script. Surface antigen genes were excluded from the analysis as the cell cultures were not controlled for serotype and surface antigen mRNA can account for 3% of total cellular mRNA. Differential gene expression was determined from fragment counts using the DESeq package[65]. *P* values were corrected for multiple testing using the Benjamini–Hochberg procedure.

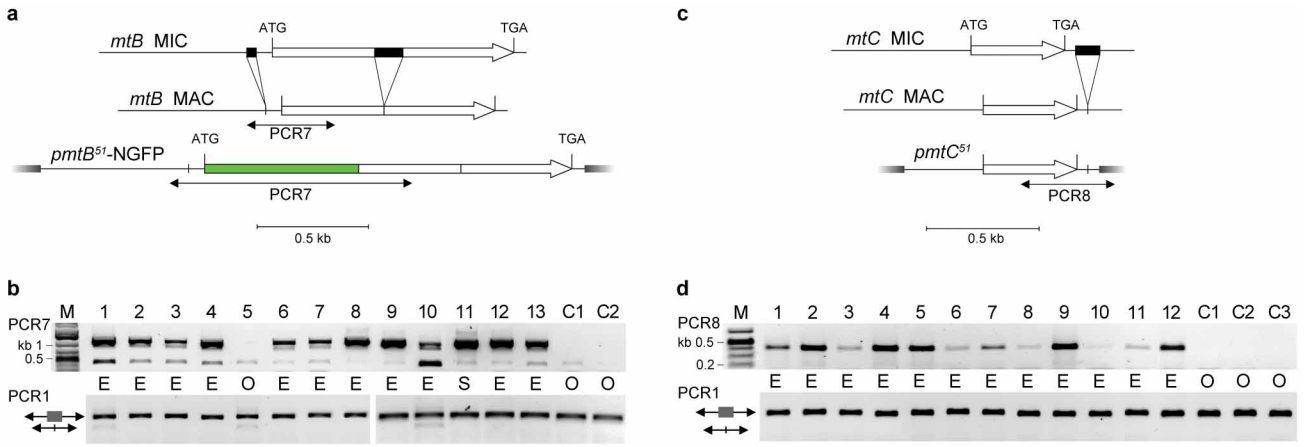
50. Beisson, J. *et al.* Mass culture of *Paramecium tetraurelia*. *Cold Spring Harb. Protoc.* **2010,** http://dx.doi.org/10.1101/pdb.prot5362 (2010).
51. Beisson, J. *et al.* Maintaining clonal *Paramecium tetraurelia* cell lines of controlled age through daily reisolation. *Cold Spring Harb. Protoc.* **2010,** http://dx.doi.org/10.1101/pdb.prot5361 (2010).
52. Pall, G. S. & Hamilton, A. J. Improved northern blot method for enhanced detection of small RNA. *Nature Protocols* **3,** 1077–1084 (2008).
53. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4,** 249–264 (2003).
54. Smyth, G. K. & Speed, T. Normalization of cDNA microarray data. *Methods* **31,** 265–273 (2003).
55. Arnaiz, O. *et al.* Gene expression in a paleopolyploid: a transcriptome resource for the ciliate *Paramecium tetraurelia*. *BMC Genom.* **11,** 547 (2010).
56. Dereeper, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36,** W465–W469 (2008).
57. Nakai, K. & Horton, P. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* **24,** 34–35 (1999).
58. Kall, L., Krogh, A. & Sonnhammer, E. L. An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics* **21** (suppl. 1), 251–257 (2005).
59. Beisson, J. *et al.* DNA microinjection into the macronucleus of *Paramecium*. *Cold Spring Harb. Protoc.* **2010,** http://dx.doi.org/10.1101/pdb.prot5364 (2010).
60. Beisson, J. *et al.* Silencing specific *Paramecium tetraurelia* genes by feeding double-stranded RNA. *Cold Spring Harb. Protoc.* **2010,** http://dx.doi.org/10.1101/pdb.prot5363 (2010).
61. Marker, S., Le Mouel, A., Meyer, E. & Simon, M. Distinct RNA-dependent RNA polymerases are required for RNAi triggered by double-stranded RNA versus truncated transgenes in *Paramecium tetraurelia*. *Nucleic Acids Res.* **38,** 4092–4107 (2010).
62. Aury, J. M. *et al.* Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444,** 171–178 (2006).
63. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
64. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
65. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).
66. Brygoo, Y., Sonneborn, T. M., Keller, A. M., Dippell, R. V. & Schneller, M. V. Genetic analysis of mating type differentiation in *Paramecium tetraurelia*. II. Role of the micronuclei in mating-type determination. *Genetics* **94,** 951–959 (1980).
67. Brygoo, Y. Genetic analysis of mating-type differentiation in *Paramecium tetraurelia*. *Genetics* **87,** 633–653 (1977).
68. Kapusta, A. *et al.* Highly precise and developmentally programmed genome assembly in *Paramecium* requires ligase IV-dependent end joining. *PLoS Genet.* **7,** e1002049 (2011).



Extended Data Figure 1 | The 195-bp segment contains the *mtA* promoter and ensures maternal inheritance of its own retention. **a**, Plasmids used for microinjection into the MAC of O cells. The thick grey lines on either side of inserts represent vector sequences. Plasmid pmtA-E contains the E MAC version with entire intergenic regions; pmtA-ES lacks the downstream unannotated gene. pmtA-IES lacks intergenic sequences upstream of the 195-bp segment. pIES contains only the 195-bp segment. The PCRs used for testing transformation are shown. **b**, Transformation of O cells with pmtA-E or pmtA-ES. PCR1 amplifies the promoter-containing form from either plasmid and the promoter-excised form from the endogenous *mtA* gene; relative amounts provide an indication of plasmid copy numbers. The mating types of injected clones and of their mass post-autogamous progeny are indicated. C1 and C2, uninjected controls; pE, control PCR on pmtA-E. **c**, Transformation of O cells with pmtA-IES. The plasmid-specific PCR2 (1,533 bp) identifies transformed clones. The O type of clones 9 and 10 could be due to transgene silencing. **d**, Transformation of O cells with pIES. Injected clones were tested with a duplex PCR: PCR4 amplifies a 450-bp product from the plasmid and PCR3 a 1,035-bp product from the endogenous *mtA* gene, hardly detectable in high-copy transformants. C1 and C2, uninjected controls; pD, control duplex PCR on an equimolar mix of pIES and pmtA-E. PCR1 revealed that the 195-bp segment was retained in the mass progeny of transformed clones; the selfer phenotype (S) of the mass progeny from clone 4 probably reflects heterogeneity among individual post-autogamous clones.



Extended Data Figure 2 | Genetic analysis of mutations affecting mating-type expression. **a**, Mendelian segregation of a pair of alleles. In both conjugation and autogamy, the two MICs undergo meiosis, but only one haploid product is retained in each cell; an additional mitosis then produces two identical gametic nuclei. During conjugation of genetically different cells (m/m and +/+), reciprocal exchange of one gametic nucleus, followed by karyogamy, therefore results in genetically identical zygotic nuclei in the two exconjugants. The drawing shows the heterozygous MICs and MACs that develop from copies of the zygotic nuclei in each of the F_1 cells. During autogamy, the two identical gametic nuclei fuse together, resulting in entirely homozygous zygotes; post-autogamous F_2 progeny of heterozygotes have a 50% probability to keep each of the parental alleles. **b**, Maternal (cytoplasmic) inheritance of mating types. Coloured Os and Es around the cells indicate the mating type expressed by each vegetative clone; the MACs are coloured to symbolize their mating-type determination states (O, blue; E, orange). Because little cytoplasm is exchanged during conjugation, the parental MACs independently condition zygotic MACs for the same mating types, resulting in cytoplasmic inheritance: the F_1 derived from the O parent is determined for O in 94% of cases, and the F_1 derived from the E parent is E in 98% of cases[66]. The frequency of mating-type reversal is much lower at autogamy: <1/50,000 in the O-to-E direction, and 1/3,000 in the E-to-O direction[67]. **c**, Cross of O-determined expression mutants mtA^O, mtB^O and mtC^O to wild-type E cells. F_2 mutant homozygotes formed in the E cytoplasmic lineage express mating type O as a result of the mutation, but remain determined for E (orange MAC). **d**, Cross of O-expressing, E-determined mutants mtA^O, mtB^O and mtC^O (as produced in **c**) to wild-type E cells. E determination of the mutant parent is revealed by E determination and expression of the derived F_1, which has received the wild-type allele. In this cross all cells are determined for E, and the expressed mating type simply depends on genotype. **e**, *mtA* promoter retention in the cross shown in **c**, illustrated here for a cross of the mtA^O mutant with wild-type E cells. Top panel: PCR analysis (PCR5) of two sets of 4 mtA^O/mtA^+ F_1 karyonides (after fertilization, two new MACs develop in each cell from copies of the zygotic nucleus, and then segregate to daughter cells, called karyonides, at the first cellular division; each pair of conjugants thus gives rise to 4 F_1 karyonidal clones). Each set contained 2 O clones that had excised the *mtA* promoter, and probably derived from the mutant O parent, and two E clones that retained it, and probably derived from the wild-type E parent (mating types are indicated below each lane). Bottom panels: after autogamy of two mtA^O/mtA^+ F_1 heterozygotes of mating type E (clone 4 in each set), 12 F_2 homozygous progeny were isolated for each, grown and tested for mating-type expression and for *mtA* promoter retention using PCR6 (Supplementary Table 6), which amplifies products of 665 bp and 470 bp from the promoter-containing and promoter-excised versions, respectively. All clones retained the promoter, although a fraction of them (14 of 23) expressed mating type O, as expected for mtA^O homozygotes. ND, not determined. **f**, *mtA* promoter retention in F_1 heterozygotes from the cross shown in **d**, illustrated by a typical set of 4 mtB^O/mtB^+ F_1 karyonides from the cross of an E-determined, O-expressing mtB^O homozygote (as produced in **c**) to wild-type E cells. PCR5 showed that the *mtA* promoter was now retained in F_1 karyonides from both parents; all were of mating type E.



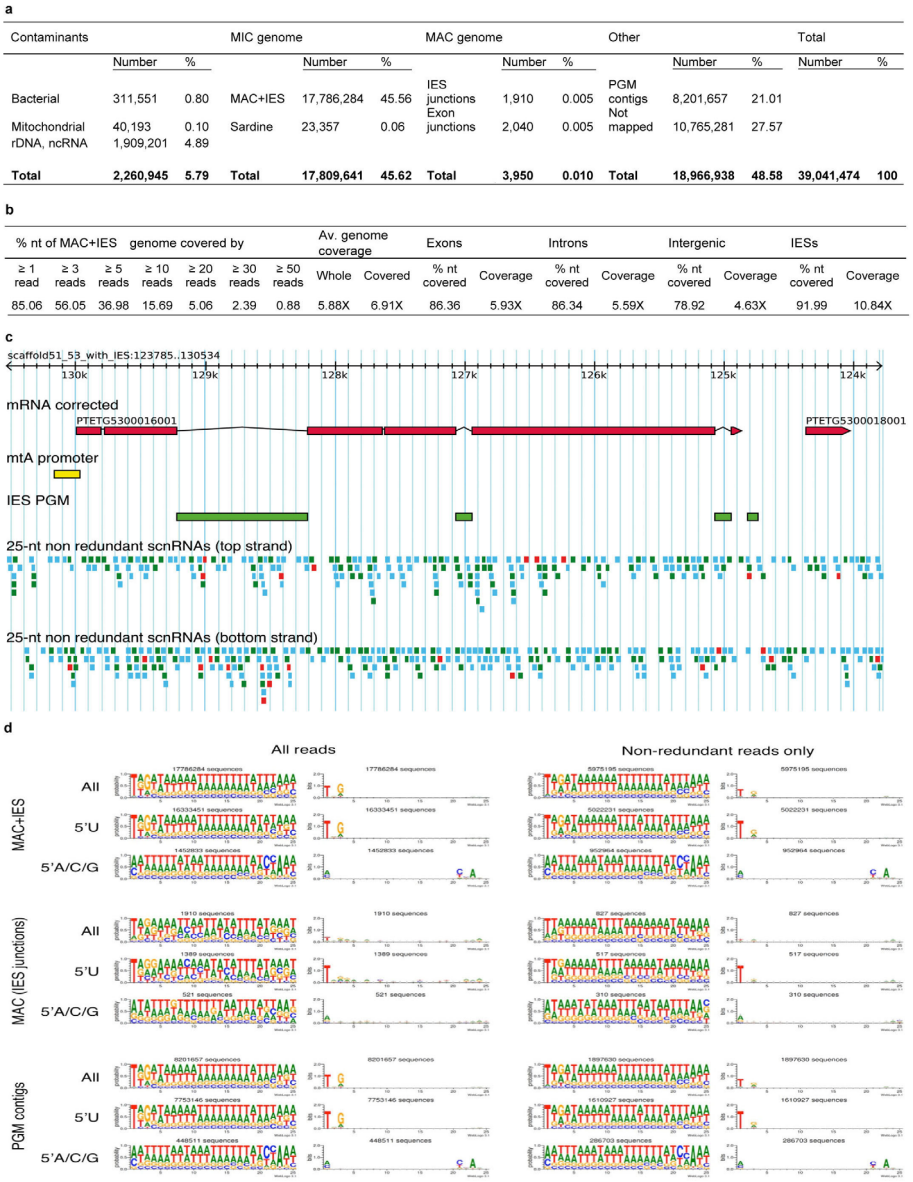
Extended Data Figure 3 | Complementation of the *mtB*^{O} and *mtC*^{O} phenotypes with the wild-type alleles of GSPATG00026812001 and GSPATG00009074001, respectively. **a**, Structure of the MIC and MAC versions of the *mtB* gene, and of the GFP fusion transgene used for complementation. The coding sequence (open arrow) is shown with the complete upstream and downstream intergenic regions. The MIC version contains two IESs (black boxes). Plasmid *pmtB*^{51}-NGFP contains the MAC version with complete intergenic regions, and the EGFP coding sequence was fused at the 5′ end of the *mtB* coding sequence. Thick grey lines on either side represent plasmid vector sequences. **b**, PCR analysis and mating types of E-determined *mtB*^{O} mutant clones transformed with *pmtB*^{51}-NGFP. PCR7 (top panel) amplifies products of 1,148 bp from the plasmid, and of 419 bp from the endogenous *mtB* gene (Supplementary Table 6). The relative abundance of the two products gives an indication of plasmid copy number in each clone. C1 and C2, uninjected control clones. Mating types are indicated below each lane. Clones containing detectable plasmid amounts expressed mating type E, indicating that the GFP fusion protein is functional; the selfing phenotype (S) of clone 11 may be due to some cells having lost the plasmid. PCR1 (bottom panel) confirmed that all clones retained the *mtA* promoter. **c**, Structure of the MIC and MAC versions of the *mtC* gene, and of the plasmid used for complementation. The coding sequence (open arrow) is shown with the complete upstream and downstream intergenic regions. The MIC version contains one IES (black box). Plasmid *pmtC*^{51} contains the MAC version with 349 bp and 98 bp of upstream and downstream intergenic sequences, respectively. The plasmid-specific PCR8 amplifies a 419-bp product (Supplementary Table 6). **d**, PCR analysis and mating types of E-determined *mtC*^{O} mutant clones transformed with *pmtC*^{51}. PCR8 (top panel) shows that all positive clones expressed mating type E. C1, C2 and C3, uninjected control clones. PCR1 (bottom panel) confirmed that all clones retained the *mtA* promoter.

a

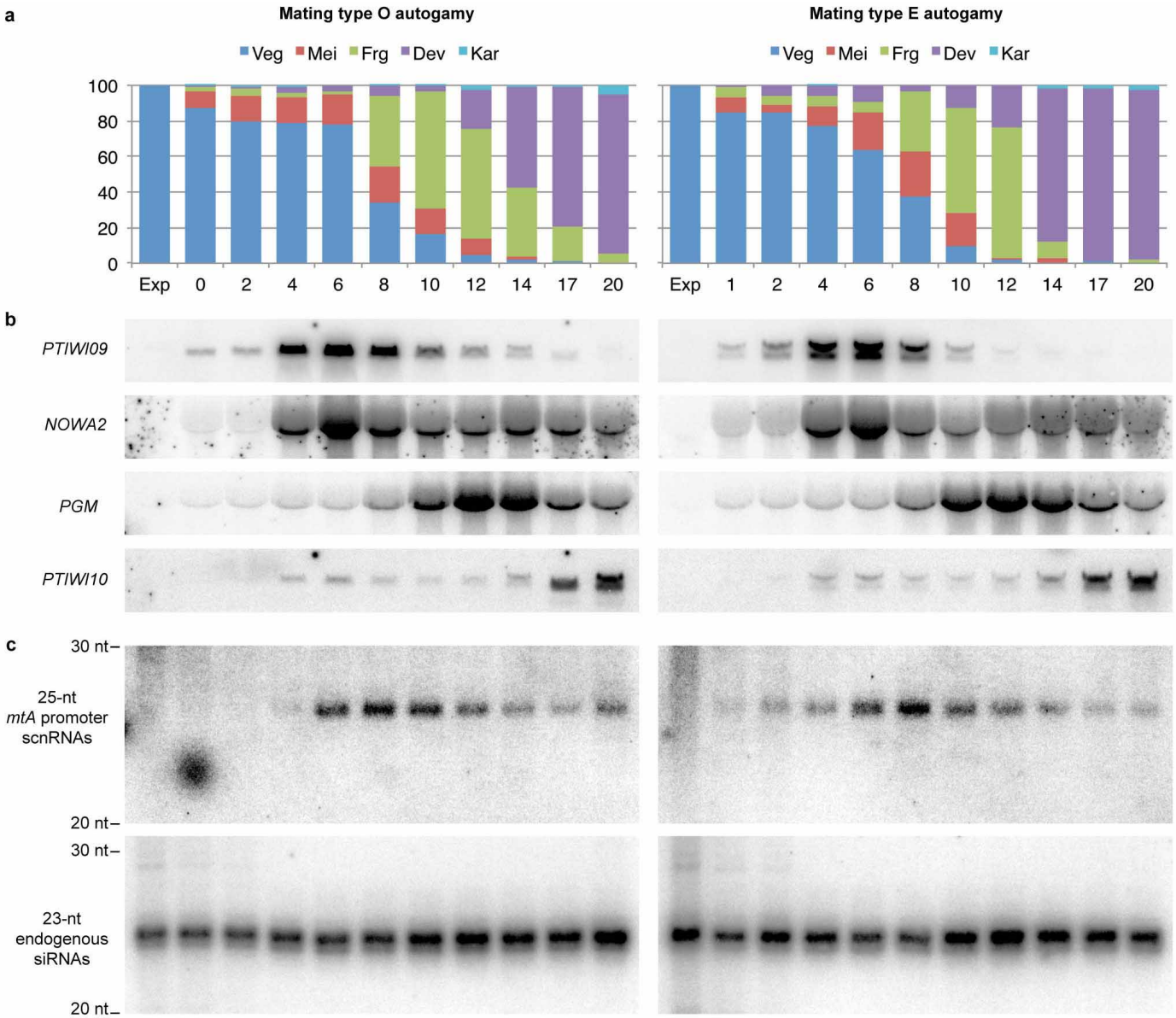
Contaminants			MIC genome			MAC genome			Other			Total	
	Number	%		Number	%		Number	%		Number	%	Number	%
Bacterial	311,551	0.80	MAC+IES	17,786,284	45.56	IES junctions	1,910	0.005	PGM contigs	8,201,657	21.01		
Mitochondrial	40,193	0.10	Sardine	23,357	0.06	Exon junctions	2,040	0.005	Not mapped	10,765,281	27.57		
rDNA, ncRNA	1,909,201	4.89											
Total	**2,260,945**	**5.79**	**Total**	**17,809,641**	**45.62**	**Total**	**3,950**	**0.010**	**Total**	**18,966,938**	**48.58**	**39,041,474**	**100**

b

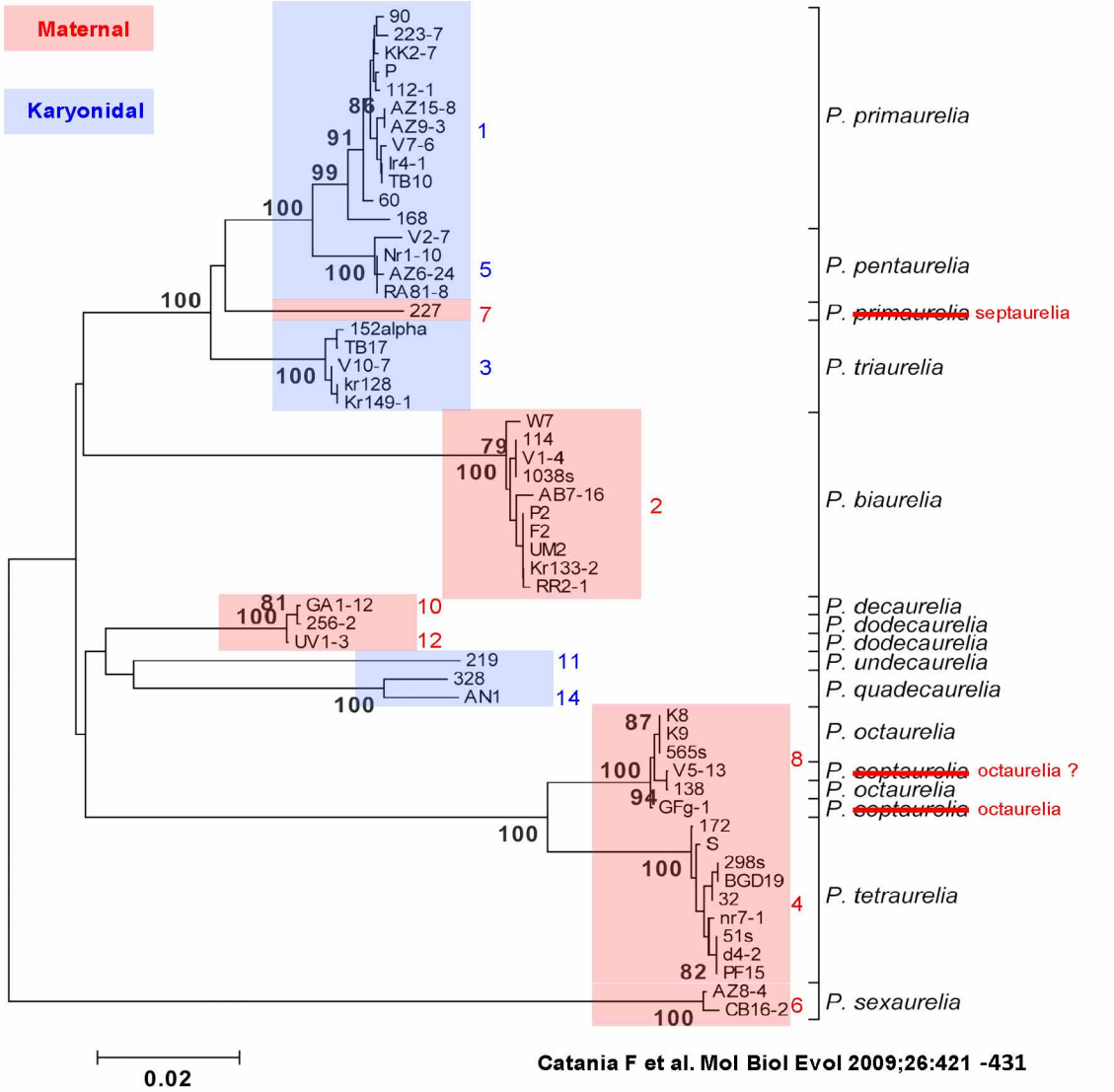
% nt of MAC+IES genome covered by							Av. genome coverage		Exons		Introns		Intergenic		IESs	
≥ 1 read	≥ 3 reads	≥ 5 reads	≥ 10 reads	≥ 20 reads	≥ 30 reads	≥ 50 reads	Whole	Covered	% nt covered	Coverage	% nt covered	Coverage	% nt covered	Coverage	% nt covered	Coverage
85.06	56.05	36.98	15.69	5.06	2.39	0.88	5.88X	6.91X	86.36	5.93X	86.34	5.59X	78.92	4.63X	91.99	10.84X



Extended Data Figure 4 | Deep sequencing of scnRNAs from an early conjugation time point (early meiosis). A total of 39,041,474 small-RNA sequences of 25 nucleotides in length were obtained by Illumina sequencing of four libraries previously constructed from gel-purified molecules migrating at 23, 24, 25, or 26 nucleotides (from the 7h*ND7* total RNA sample[12]). **a**, 25-nucleotide reads were first mapped (no mismatch allowed) on possible contaminant sequences (genomes of bacteria commonly found in cultures, *P. tetraurelia* mitochondrial genome, *P. tetraurelia* rDNA and other non-coding RNAs). The remainder was then mapped on known MIC sequences (the 'MAC+IES' genome, and 9 individual copies of the Sardine transposon[3]). To determine whether any of the remaining reads could correspond to IES excision junctions or to spliced transcripts, they were then mapped on the MAC genome and on the genome-wide set of spliced transcripts. Very few hits were found and these did not show the characteristic 5′-UNG signature of scnRNAs (see **d**), suggesting that these molecules represent longer forms of endogenous siRNAs and/or could be mapped because of IES or intron annotation errors. Of the remaining unmapped reads (~49%), close to one-half could be mapped on 'PGM contigs', a ~25-Mb preliminary assembly of MIC-specific sequences that are not collinear to MAC chromosomes[3] and are thus likely to represent bona fide scnRNAs. **b**, Statistics about the coverage of the 'MAC+IES' genome (17,786,284 reads). The average coverage is similar for exons, introns and intergenic regions, but is ~2-fold higher for IES sequences. This may mean either that scnRNAs are initially produced in higher amounts from IESs, or that active degradation of scnRNAs homologous to MAC sequences is already under way at this early stage. **c**, Mapping of 25-nucleotide reads on the MIC version of the *mtA* gene region. Coding sequences of the corrected *mtA* gene model (PTETG5300016001, after correction of assembly indels and re-annotation) and of the short gene downstream (PTETG5300018001) are shown as red boxes interrupted by introns and IESs. The yellow box represents the 195-bp segment of the *mtA* promoter that is excised in O cells; the 4 IESs are shown as green boxes. 25-nucleotide reads mapping on the top or bottom strands of the region are colour-coded to indicate the number of times each one was sequenced: blue, one read; green, 2–9 reads; red, ≥10 reads. **d**, Compositional profiles (nucleotide frequency on the left, and deviation from randomness on the right) of reads mapping to the MAC+IES genome, to the MAC genome, or to 'PGM contigs'. For each set, logos are shown for all reads (left), or for the non-redundant subset only (right). For the 'MAC+IES' and 'PGM contigs' sets, the logos computed from all reads clearly show the 5′-UNG signature typical of scnRNA guide strands, as is the case for the major subset of reads starting with U (5′U), while the minor subset of sequences not starting with U (5′A/C/G), which may represent the steady-state amount of passenger strands, shows the complementary signature CNA at positions 21–23. Deviation from randomness is greater when computed from all reads than when computed from the non-redundant subset only, indicating that molecules with the signature are intrinsically produced in higher amounts, or are more stable. The small set of reads that mapped only to the MAC genome (putative IES excision junctions) does not show a clear 5′-UNG signature, suggesting that most of those are not scnRNAs. The same is true of the small number of reads mapping to exon–exon junctions (not shown).

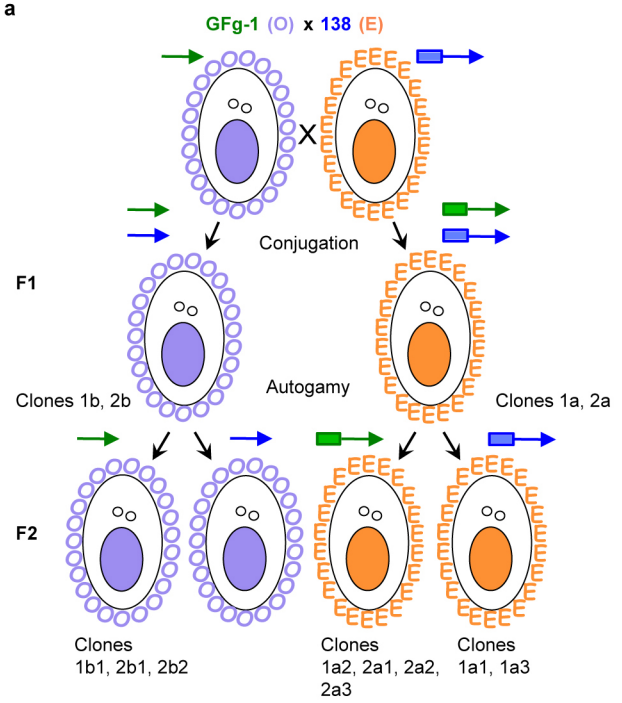


Extended Data Figure 5 | Northern blot analysis of *mtA*-promoter scnRNAs during autogamy of O or E cells. Mass cultures were allowed to starve, and RNA samples were extracted during exponential growth (Exp) and then at different times (0–20 h) after the appearance of the first meiotic cells. Cells become committed for autogamy at a fixed point of the cell cycle, so that the best synchrony that can theoretically be achieved is the duration of one cell cycle; in these experiments, the time between the first and the last cells to begin meiosis was ~12 h. **a**, Proportions of cells in different cytological stages at each time point, as determined by DAPI staining. Veg, vegetative cells; Mei, meiosis (crescent stage, meiosis I, meiosis II); Frg, cells with fragmented old MAC but new MACs not yet clearly visible; Dev, cells with two clearly visible developing new MACs; Kar, cells after the karyonidal division, with only one developing new MAC. **b**, Northern blot analysis of mRNAs for early (*PTIWI09*, *NOWA2*), middle (*PGM*), or late (*PTIWI10*) genes. The same blots were hybridized successively with the 4 probes. **c**, Northern blot analysis of small RNAs. The top panels show hybridization with the 195-bp *mtA*-promoter probe, revealing accumulation of 25-nucleotide scnRNAs slightly later than expression of the meiosis-specific genes *PTIWI09* and *NOWA2*. As a control, the same blots were rehybridized with an oligonucleotide probe specific for a cluster of 23-nucleotide endogenous siRNAs on scaffold 22[12,61], which are abundantly produced at all stages of the life cycle (bottom panels). Quantification of the *mtA*-promoter scnRNA signal and normalization with the siRNA signal did not reveal any significant difference in their amount or timing between the two mating types (not shown). Previous studies showed that the double-strand breaks that initiate IES excision in the new MACs start being detectable before the maximum of expression of the putative endonuclease *PGM*[5,68] (no later than 10 h in these time courses). A PCR analysis of post-autogamous DNA samples confirmed that the *mtA* promoter was fully excised in mating type O, and fully maintained in mating type E (not shown).



Extended Data Figure 6 | Phylogenetic tree of different strains from most *P. aurelia* species based on sequence polymorphisms in three nuclear genes. This figure is modified with permission from figure 2 of Catania *et al.* Genetic diversity in the *Paramecium aurelia* species complex *Mol. Biol. Evol.* 2009, 26, 421–431 (ref. 32). Wrong species assignment of some strains studied here are corrected in red. Strain V5-13 was probably mis-assigned to *P. septaurelia* through the same error as for strain GFg-1 (see Extended Data Fig. 7). Species showing maternal inheritance or random determination of mating types are highlighted in red and blue, respectively.

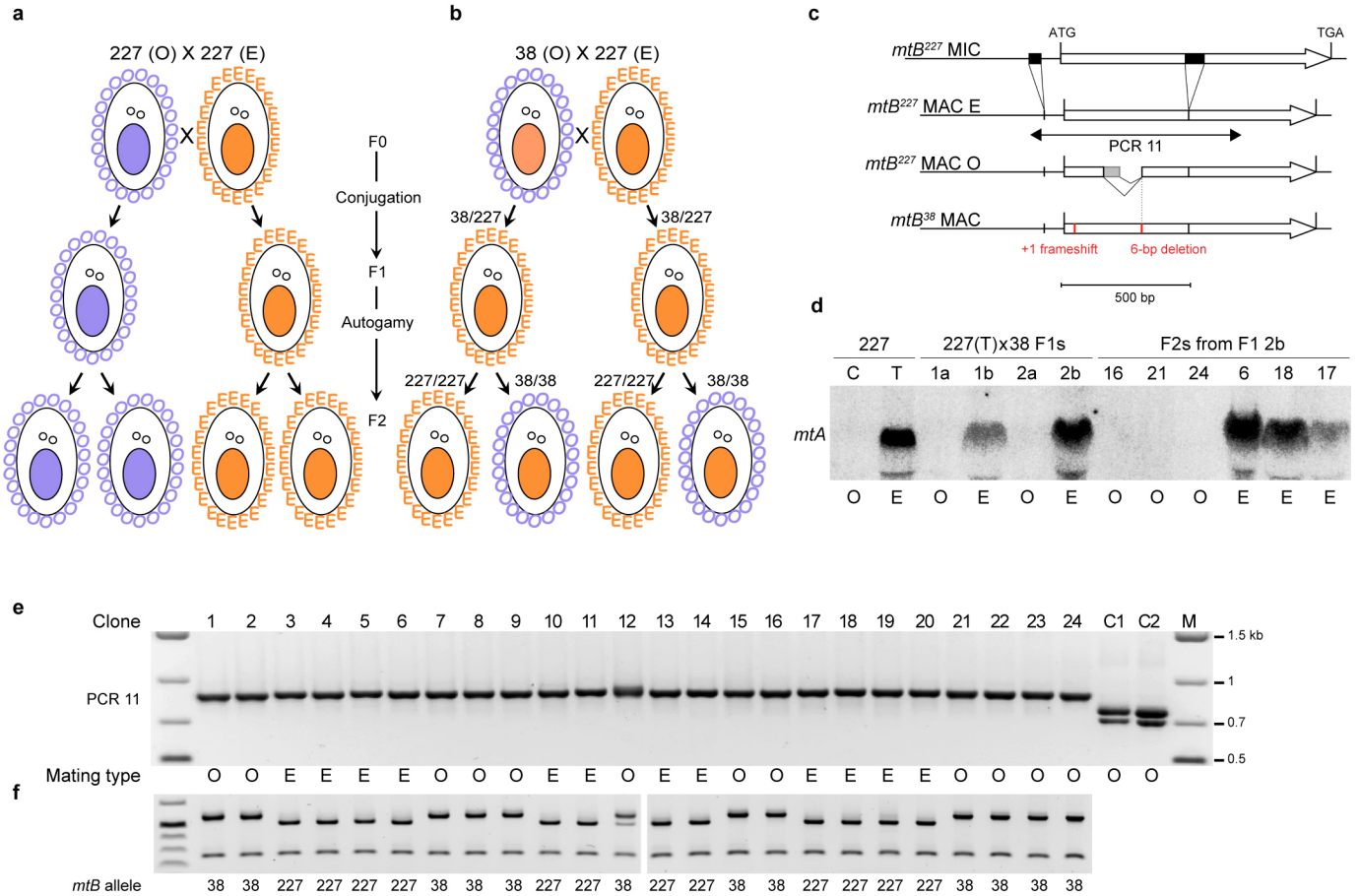
a



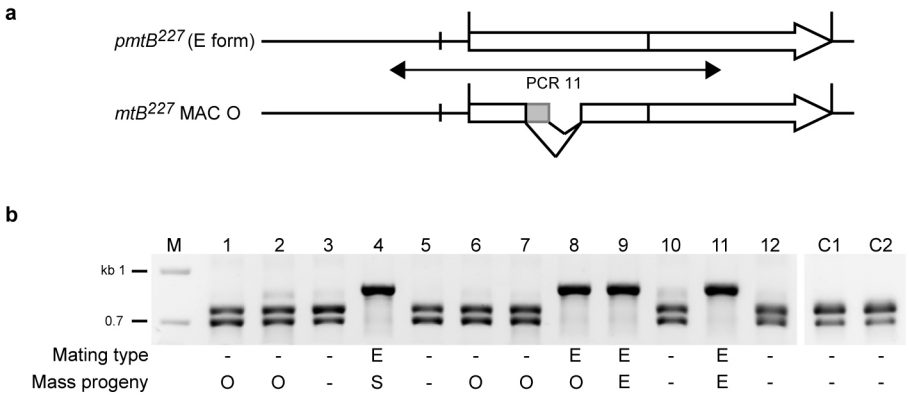
b

	Mitochondrial COXII variant	*mtA* genotype	*mtA* promoter status	Inferred mating type
F1 pairs				
1a	138	ND	ND	E
1b	GFg-1	138/GFg-1	excised	O
2a	138	ND	ND	E
2b	GFg-1	138/GFg-1	excised	O
F2 clones, E side				
1a1	138	138/138	retained	E
1a2	138	GFg-1/GFg-1	retained	E
1a3	138	138/138	retained	E
2a1	138	GFg-1/GFg-1	retained	E
2a2	138	GFg-1/GFg-1	retained	E
2a3	138	GFg-1/GFg-1	retained	E
F2 clones, O side				
1b1	GFg-1	GFg-1/GFg-1	excised	O
2b1	GFg-1	GFg-1/GFg-1	excised	O
2b2	GFg-1	GFg-1/GFg-1	excised	O

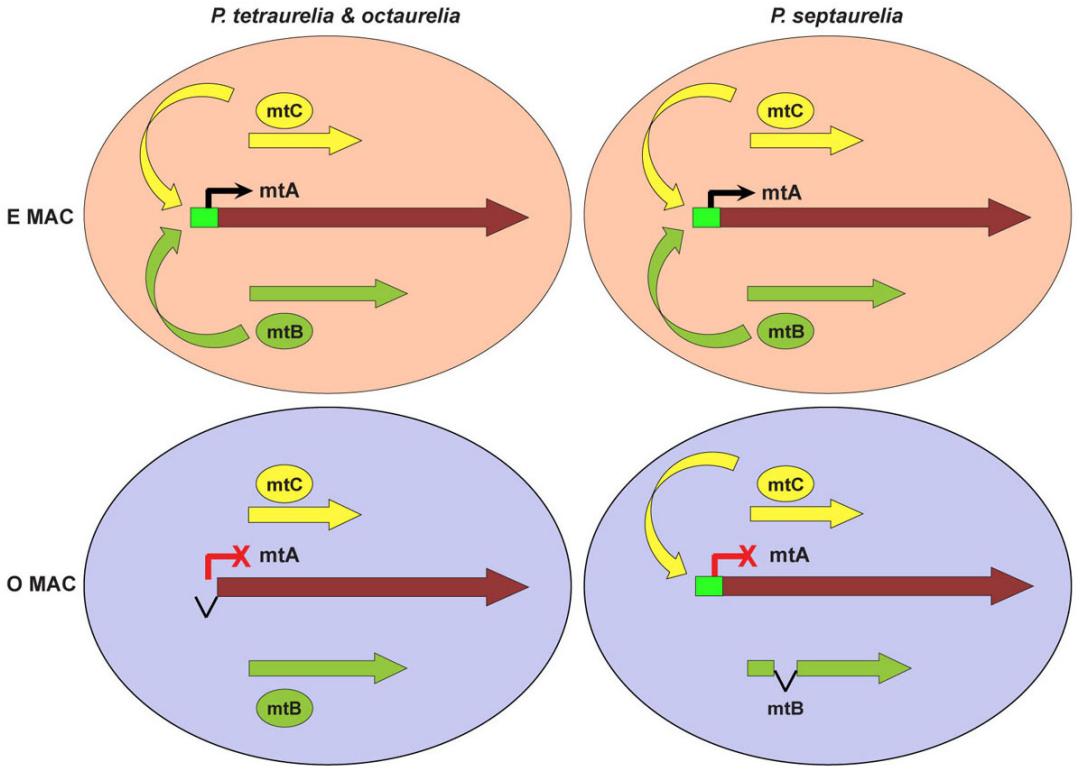
Extended Data Figure 7 | Strain GFg-1 belongs to the same species as strain 138, that is, *P. octaurelia*. GFg-1 is among a set of strains that were originally assigned to *P. septaurelia* on the basis of conjugation tests with strain 38, a reference strain for that species. However, the stock of strain 38 used in these tests was not 38, but instead some *P. octaurelia* strain, as shown by the comparison of *mtA* sequences with those from the original strain 38 obtained from ATCC (ATCC number 30575) and those from strain 138, a reference strain for *P. octaurelia*. **a**, Scheme of the cross GFg-1 (O) × 138 (E). The mating types of parents were determined by cross-agglutination with *P. tetraurelia* tester lines. The green and blue arrows beside each cell represent the *mtA* gene, colour-coded to indicate the GFg-1 and 138 alleles; the box at the 5′ end symbolizes retention of the *mtA* promoter in the MAC genome on the E side of the cross. **b**, Molecular characterization of two pairs of F_1 heterozygotes, and of some F_2 clones obtained by autogamy of these F_1 heterozygotes. The parental origin of each F_1 clone was ascertained by a sequence polymorphism in the mitochondrial *COXII* gene (PCR9, Supplementary Table 6). PCR amplification and sequencing of a segment of the *mtA* gene encompassing the promoter (PCR10, Supplementary Table 6) showed that the two F_1 clones deriving from the GFg-1 parent were heterozygotes, and that the *mtA* promoter was precisely excised from both alleles (see Supplementary Data 1a). ND, not determined. Analysis of 6 viable F_2 clones obtained by autogamy of the F_1 clones deriving from the 138 parent showed that 4 of them were homozygous for the GFg-1 allele, whereas the other two were homozygous for the 138 allele; the *mtA* promoter was retained in all cases. The evidence for successful genetic exchange between strains GFg-1 and 138 and for viable recombinant F_2 progeny demonstrates that these strains belong to the same species; that is, *P. octaurelia*.



Extended Data Figure 8 | Genetic and molecular analysis of the cross between strains 227 and 38 of *P. septaurelia*. **a**, Maternal inheritance of mating types in the wild-type strain 227. **b**, Strain 38 is genetically restricted to O expression, but constitutively determined for E. The mutation identified in strain 38 is of particular interest because it affects both the expression and the inheritance of mating types, which suggests that it lies in a gene that controls mating-type determination through an alternative rearrangement. Indeed, known *P. tetraurelia* mutations fall in two distinct categories. mtA^O, mtB^O and mtC^O prevent expression of type E but have no effect on the rearrangement that determines mating types or on its maternal inheritance, whereas mtF^E lies in a *trans*-acting factor required for a subset of rearrangements during MAC development but has no effect on the expression of mating types during sexual reactivity. The only type of mutation that can be envisioned to affect both expression and determination/inheritance would be a mutation preventing expression of a functional protein required for E expression, and at the same time preventing in *cis* (by destroying a potential Pgm cleavage site) the rearrangement that normally inactivates this gene in the MAC of wild-type O cells. The mt^{XIII} allele of strain 38 restricts cells to O expression in a recessive manner, but also has a maternal effect that enforces constitutive E determination in sexual progeny. Notably, elegant experiments showed the latter effect to be dominant[20]: the sexual progeny of a cell carrying at least one mt^{XIII} allele can never be determined for O or transmit O determination. **c**, Sequencing of the mtB^{38} allele revealed features that may account for both effects. A frameshift mutation makes it a pseudogene, explaining the genetic restriction to O expression. In addition, a 6-bp deletion removes one of the IES-like boundaries used in the mtB^{227} allele for coding-sequence deletions in O clones. Given the requirements of IES excision in *P. tetraurelia* and in sibling species[36], this can be predicted to make the same deletions impossible in the mtB^{38} allele, which would explain the constitutive E determination effect. The full-length mtB^{38} pseudogene in the MAC of 38 cells would indeed be expected, after a cross to 227, to protect the highly similar zygotic mtB^{227} allele against coding-sequence deletions in the derived F_1, through titration of homologous scnRNAs. **d**, Molecular analysis of the 38 × 227 cross. To verify this maternal effect, we crossed an E-expressing 227 clone ($pmtB^{51}$-transformed clone T, same as in Fig. 5b; C, uninjected control) with strain 38, and F_1 heterozygotes were tested for mating types and for *mtA* expression by northern blotting. As expected, F_1 heterozygotes deriving from the 38 parent (1b and 2b, as determined by sequencing of a mitochondrial polymorphism) were E and expressed *mtA*, indicating that the incoming mtB^{227} allele had been rearranged into a functional, full-length form in the MAC. After autogamy of F_1 clone 2b, 24 independent F_2 homozygotes were isolated and tested for mating types. Consistent with the Mendelian segregation of *mtB* alleles, 12 were O and 12 were E (Supplementary Table 5); northern blot analysis of 3 clones of each type showed that only E clones expressed *mtA*. **e**, All F_2 clones maintained the full-length *mtB* gene in the MAC. PCR11 (Supplementary Table 6) amplifies an 888-bp fragment from the MAC version of mtB^{38}, and an 893-bp fragment from the full-length MAC version of mtB^{227}. C1 and C2, control PCR11 on two O clones of strain 227, showing the 806-bp and 744-bp fragments resulting from the two alternative coding-sequence deletions. **f**, Mating types co-segregate with *mtB* alleles among F_2 homozygotes. Digestion of the PCR products with AluI distinguishes the 38 and 227 alleles. *mtA* and *mtC* alleles segregated independently (Supplementary Table 5).



Extended Data Figure 9 | Transformation of *P. septaurelia* 227 O cells with the full-length *mtB*^{227} gene induces heritable E expression. **a**, Plasmid *pmtB*^{227}, containing the full-length E MAC form of *mtB*^{227}, was microinjected into the MACs of O cells of strain 227, and 12 injected clones were tested by PCR11. **b**, This PCR amplifies an 893-bp product from the transgene, and products of 744 and 806 bp from the two internally deleted MAC forms of the endogenous *mtB*^{227}. C1 and C2, control uninjected clones. Clones 4, 8, 9 and 11 contained high copy numbers of the transgene, so that the PCR products from the endogenous gene are not detectable. Because only O tester lines were available, the mating types of injected clones could be ascertained only for E-expressing clones, as indicated below each lane. This was the case of the 4 high-copy clones; other clones were presumably O, although it was impossible to make sure that the cells were sexually reactive (indicated by '-'). Injected clones were then taken through autogamy, and the mass progenies were again tested for mating types. In this case, the mass progenies from clones 9 and 11, which proved to be pure E, were used as E testers to carry out a full test for some of the other progenies. Of the other two transformed clones expressing E, clone 4 gave rise to a selfing progeny (S), probably a mix of O and E clones, whereas clone 8 gave rise to a pure O progeny.



Extended Data Figure 10 | A general model for mating-type determination in *P. aurelia* species. In the three species examined, mating type E depends on expression of the mtA protein during sexual reactivity. *mtA* transcription in turn requires the *mtB* and *mtC* gene products (the requirement for *mtC* in *P. septaurelia*, and for both genes in *P. octaurelia*, remains to be verified). In *P. tetraurelia* and *P. octaurelia*, mating type O is determined during MAC development by excision of the *mtA* promoter as an IES, preventing expression of the gene. In *P. septaurelia*, mating type O is determined by the excision of segments of the *mtB* coding sequence as IESs, which similarly prevents *mtA* expression.