Breaking the species barrier: *recurrent genomic introgressions from very distant lineages in a ciliate*

Florian Bénitière (1), Olivier Arnaiz (2), Simon Penel (1), Sandra Duharcourt (3), Eric Meyer (4), Linda Sperling (2), Laurent Duret* (1)

(1) Universite Claude Bernard Lyon 1, LBBE, UMR 5558, CNRS, VAS, Villeurbanne, France (2) Université Paris-Saclay, CEA, CNRS - Institute for Integrative Biology of the Cell (I2BC), Gif-sur-Yvette, France (3) Université Paris Cité, CNRS, Institut Jacques Monod, F-75013 Paris, France (4) Institut de Biologie de l'ENS (IBENS), Département de Biologie, École Normale Supérieure, CNRS, INSERM, Université PSL, Paris, France
(*) corresponding author

Keywords
Speciation; species barriers; hybridization; ciliates; paramecium

Abstract
The process of speciation, through which two reproductively isolated species emerge from an ancestral population, is often gradual, including a period during which gene flow still occurs, and hence speciation remains reversible. However, if isolation persists, the progressive accumulation of genetic incompatibilities ultimately locks the speciation process, reaching the point where hybrid offspring are inviable or sterile. In plants and animals, the establishment of species barriers is quite rapid: beyond a net genetic divergence of 0.02 substitutions per synonymous site, interspecific introgressions become extremely rare. Here we report that in ciliates, species barriers can be broken despite huge genetic divergence. We show that the genome of *Paramecium sonneborni* contains hundreds of DNA segments that were acquired by recurrent horizontal transfers from multiple very distant lineages (synonymous divergence ~1 substitution/site). These segments approach the size of full-length chromosomes, which indicates that they were introgressed by inter-specific mating. These data show that in the ancestry of *P. sonneborni*, hybridization events with highly divergent species occurred repeatedly and led to viable and fertile offspring. We propose that this unexpected promiscuous sexual behavior is made possible by the fact that paramecia separate somatic and germline functions in two types of nuclei: the elimination of non-self DNA fragments during the development of the somatic genome allows individuals of a species to remain phenotypically the same, despite recurrent genetic flow from distant species into their germline.

Introduction

An important issue in evolutionary biology is to understand the process of speciation, i.e. how two reproductively isolated species emerge from an ancestral population. Various factors can initiate the speciation process (e.g. isolation by geographic barriers, adaptation to distinct ecological niches, etc.), but whatever the initial factor, once two populations become reproductively isolated from each other, they start evolving independently, and hence progressively accumulate genetic incompatibilities. These so-called Bateson–Dobzhansky–Muller incompatibilities (BDMIs) decrease the fitness of hybrid individuals, and ultimately lead to their sterility or non-viability (Orr 1996). The speciation process is often gradual, i.e. includes a 'grey zone' during which genetic introgressions are still possible between "semi-isolated" species (Stankowski and Ravinet 2021). Thus, if secondary contacts occur during the

early phase, then hybridizations may restore genetic homogeneity, thereby stopping the speciation process. However, as BDMIs accrue, selection is expected to favor the evolution of traits that prevent interspecific mating, thereby further reinforcing genetic isolation. A key point is therefore to unravel the factors that determine the rate at which BDMIs accumulate. The analysis of population genomics data from a wide range of taxa has shown that introgressions are restricted to species that are genetically very close: in animals, gene flow appears to be abolished as soon as the net synonymous divergence exceeds 0.02 substitutions per site, and this threshold is even lower in plants (Roux et al. 2016; Monnet et al. 2025). Thus, although hybridization is quite common in nature, it generally leads to inviable or sterile offspring, and hence does not contribute to gene flow, except between very closely related species (Edelman and Mallet 2021).

Here we report an extreme exception to this rule, that we discovered incidentally in the course of sequencing the genomes of several paramecia species (Arnaiz et al. 2025). We present evidence of recurrent events of introgression within the genome of *Paramecium sonneborni*, originating from very remote species (synonymous divergence ~1 substitution/site), and we propose a model explaining how fertile hybrids can be formed between such distant species, despite BDMIs.

To facilitate the interpretation of our observations, it is useful to give some background information on the biology and genetics of *Paramecium* species. These are aquatic unicellular eukaryotes that are commonly found in ponds, where they feed on bacteria, yeast, or algae. They were among the first ciliates to be observed, thanks to the invention of the microscope in the late 17th century, and since then have become a widely used model for biological studies (Beisson et al. 2010). Their life cycle includes a vegetative phase (reproduction by mitotic cell divisions), followed by a sexual phase, which involves either conjugation (the pairing and reciprocal fertilization of two cells of complementary mating types) or autogamy (a self-fertilization process that evolved in an ancestor of the *P. aurelia* clade) (Sonneborn 1937).
Like all ciliates, *Paramecium* cells possess two types of nuclei: a highly polyploid somatic macronucleus (MAC), responsible for gene expression, and a diploid germline micronucleus (MIC), which is not expressed, but where meiosis occurs. At each sexual generation, the parental MAC is lost, while new MIC and MAC develop from copies of the zygotic nucleus. During MAC development, many DNA segments are reproducibly eliminated, thus leading to a somatic genome that is smaller than the germline genome (Bétermier et al. 2023; Balan et al. 2024).
The genus *Paramecium* encompasses a wide diversity of species. Based on cytological observations, 19 species can be distinguished (De Souza et al. 2020). However, genetic crosses have revealed a large amount of cryptic diversity within these groups. Notably, early studies have established that *P. aurelia*, initially considered as one 'species', was in fact a complex of 14 sibling species ('syngens') that are morphologically indistinguishable but sexually incompatible (Sonneborn 1975). More recently, phylogenetic and population genetic studies based on MAC genome sequences have confirmed that these 14 syngens correspond to *bona fide* species, reproductively isolated and genetically divergent (for a review, see (Long et al. 2023)). Currently, 20 species have been ascribed to the *P. aurelia* species complex (Sawka-Gądek et al. 2021): 15 that are morphologically indistinguishable (named *P. primaurelia* to *P. quindecaurelia*), together with *P. sonneborni*, *P. schewiakoffi*, and three sibling species of *P. jenningsi* (Sonneborn 1975; Aufderheide et al. 1983; Fokin et al. 2004; PrzyboŚ and Tarcz 2019; Potekhin and Mayén-Estrada 2020). This clade, whose phylogeny is now well established, encompasses a wide variety of species, genetically as distant as the most remotely related mammals (Sawka-Gądek et al. 2021; Sellis et al. 2021).

Little is known regarding the ecological factors that may have driven the diversification of this clade. Some species are preferentially found in tropical environments (e.g. *P. sexaurelia, P. octaurelia, P. tredecaurelia, P. quadecaurelia,* and *P. sonneborni*), or in temperate or cold climate zones (e.g. *P. biaurelia* and *P. novaurelia*), whereas others are totally cosmopolitan (e.g. *P. primaurelia* or *P. tetraurelia*) (Melekhin et al. 2024). Overall, most species have a very widespread geographical distribution and are found in sympatry on several continents (Melekhin et al. 2024). In the *aurelia* clade, the level of neutral polymorphism ranges from 0.6% in *P. tetraurelia* to 2.7% in *P. sexaurelia* and shows no clear pattern of spatial structure, no isolation by distance, indicating recent dispersal (Johri et al. 2017). One noteworthy observation is that three rounds of whole genome duplications (WGDs) occurred in the ancestry of the *aurelia* lineage, and the last one shortly predates the radiation of the complex (Aury et al. 2006; McGrath et al. 2014; Gout et al. 2023). The differential loss of paralogous gene copies after a WGD is expected to be an important source of BDMIs (Aury et al. 2006; Scannell et al. 2006), and hence it has been suggested that these WGDs might have contributed to the diversification of the *P. aurelia* complex (Aury et al. 2006; McGrath et al. 2014; Gout et al. 2023).

To investigate the organization and evolution of the germline genome, we sequenced the MIC of 7 *P. aurelia* species (Sellis et al. 2021; Arnaiz et al. 2025). We obtained a high-quality assembly for the *P. tetraurelia* MIC genome (104 Mb), which consists of ~160 tiny chromosomes (300 kb – 1.2 Mb)(Arnaiz et al. 2025). Three genomic compartments can be distinguished, depending on their fate during the development of the new polyploid MAC: i) MIC-limited regions, that are eliminated from the MAC, ii) constitutive MAC-destined regions, that are systematically retained in the MAC, and iii) MAC-variable regions, which correspond to DNA sequences that are retained in only a small fraction of MAC copies (Sellis et al. 2021). These three compartments represent respectively 20%, 71% and 9% of the MIC genome. The MAC-destined regions are very gene-rich (~500 genes per Mb), whereas most repeated sequences are located in the MIC-limited regions (Sellis et al. 2021; Arnaiz et al. 2025). Two types of MIC-limited regions can be distinguished: the 'Internal Eliminated Sequences' (IESs, totalizing 3.9 Mb) that correspond to short sequences (typically <150bp) interspersed within MAC-destined regions and notably within genes, whereas the other eliminated sequences (16 Mb) are generally located towards the end of MIC chromosomes (Arnaiz et al. 2025). Most other species show similar genome content, with a MIC genome ~100 Mb. The only exception is *P. sonneborni*, where the MIC reaches 217 Mb (Arnaiz et al. 2025). The size of the MAC-destined compartment is similar across species, ranging from 65 Mb in *P. tredecaurelia* to 82 Mb in *P. sonneborni* (Sellis et al. 2021). Thus, the large size of the *P. sonneborni* MIC results from a huge inflation of its MIC-specific compartment.

To investigate the cause of this increase, we compared *P. sonneborni* MIC sequences to other *Paramecium* genomes, and discovered that *P. sonneborni* contained long regions with very high sequence similarity to distantly related *aurelia* species. This prompted us to systematically investigate patterns of horizontal transfer across paramecia. We found evidence of recurrent events of introgression of chromosome-size segments within the MIC-specific genomic compartment of *P. sonneborni*, originating from very distant species. We propose that these introgressions result from allotetraploidization events, and we discuss the mechanism that could explain how these events can be tolerated despite the accumulation of BDMIs.

RESULTS

Evidence of massive horizontal transfer in the exceptionally large MIC genome of P. sonneborni

To investigate the origin of the large MIC-limited compartment in *P. sonneborni*, we compared its sequences to all available genomes from 13 species of the *P. aurelia* complex (7 species for which both MIC and MAC genomes have been sequenced, and 6 species for which only the MAC is available). *P. sonneborni* is closely related to *P. jenningsi* (genetic distances at synonymous sites: dS=0.27), but quite distant from all the other species that we analyzed (Figure 1A), with dS values ranging from 0.8 to 1.0 substitution per site (Sellis et al. 2021). This implies that outside of coding regions, sequences are generally too divergent to be aligned across these species. Yet, we noticed that several *P. sonneborni* MIC contigs contain long segments with very high similarity to distantly related species. For instance, we observed a *P. sonneborni* contig of 100kb presenting long stretches above 99% identity with a MAC scaffold from *P. quadecaurelia*, a level of similarity higher than with any other *aurelia* species, including *P. jenningsi* (Figure 1B). This suggests that this long genomic segment has been subjected to horizontal transfer (HT) between the *P. sonneborni* and *P. quadecaurelia* lineages.

This unexpected observation led us to systematically analyze all MIC genomes (from 7 species) to search for loci presenting evidence of recent HT among the *aurelia* complex. For this, we divided each MIC genome (query) into non-overlapping 500 nt windows and compared (with blastn) each window with all other available *Paramecium* genomes (targets) (Figure 1A). We then searched for sets of neighboring windows (at least 2 kb-long in total) with conserved order and orientation between the query and the target, and showing a strong signal of HT (i.e. a level of sequence similarity significantly higher with a genome that is not the sister group of the focal species, than with any other *aurelia* species).

In most species we identified only a limited number of HT segments (1 to 22 segments, totalizing up to 144 kb; Figure 1C). Strikingly, in the *P. sonneborni* MIC genome, we identified 521 HT segments totaling almost 10 Mb. HT segments are on average 2 to 5 times longer in *P. sonneborni* (mean length = 18.4 kb) than in other species (Suppl. Fig. 1). In *P. sonneborni* we detected 120 HT segments longer than 25 kb (totaling 5 Mb). The length of HT segments in *P. sonneborni* is probably underestimated because the MIC assembly is quite fragmented (N50 ~33 kb). Indeed, we found several cases where a single region in the MAC genome of a target species showed synteny with multiple *P. sonneborni* contigs (Figure 1B). These observations indicate that many HT segments in *P. sonneborni* are very long, approaching the size of an entire chromosome (>220 kb for the example shown in Figure 1B). Conversely, none of the HT segments in the other species were larger than 25 kb.

Several observations allowed us to exclude the hypothesis of a DNA contamination artefact. First, contamination with DNA from the target genomes should lead to nearly identical sequences, whereas for many HT segments, the level of sequence divergence ranges between 1% and 12% (Figure 2), far too high to result from sequencing errors. Second, two of the target genomes with which we detected many HT segments (*P. quadecaurelia* and *P. novaurelia;* Figure 1A) were sequenced in the USA (Gout et al. 2023), and were never handled in the laboratory where *P. sonneborni* MAC and MIC DNA were sequenced (Genoscope, France). Finally, the sequencing depth along *P. sonneborni* HT segments is the same as that of the rest of the genome, which would not be expected if they resulted from contaminant DNA (Suppl. Fig. 2).

The HT segments detected in *P. sonneborni* involve four distinct lineages (represented by *P. sexaurelia, P. tredecaurelia, P. quadecaurelia,* and *P. novaurelia*) (Figure 1A). Among HT segments detected by comparison between *P. sonneborni* and *P. sexaurelia* (the only species pair for which we had both MIC genome assemblies), we observed many conserved IESs (Suppl. Fig. 3), which demonstrates that these HT events involved the transfer of MIC DNA and not of MAC DNA (where IESs are absent). *P. sexaurelia* belongs to the same subclade as *P. sonneborni* but is nevertheless very distant (average dS=0.8; (Sellis et al. 2021)). The three other lineages belong to subclade A, very divergent from *P. sonneborni* (dS~1; (Sellis et al. 2021)). We did not detect any HT segment with the other lineages of subclade A (Figure 1A). It should be noted that MIC genome sequences are available for only 7 of the 13 analyzed species. Because of this sampling limitation, HT segments present in the MIC genome of the query were detectable only if they encompassed MAC-destined regions in the target genome (see Methods for details). Hence, the set of segments that we identified is probably an underestimate of the total amount of HT.

P. sonneborni is the recipient of recurrent horizontal transfers from distantly related lineages

The above results show that gene flow occurred between *P. sonneborni* and several other *Paramecium* lineages. To establish the direction of these transfers, we analyzed the distribution of HT segments in the MIC of *P. sonneborni*: 85% are located in the MIC-limited compartment of *P. sonneborni*, 14% in MAC-variable regions and only 1% in MAC-destined regions (these compartments represent respectively 55%, 7% and 38% of the MIC genome) (Figure 1A). Thus, the majority of HT segments are confined to genomic compartments that are either eliminated or variably retained during MAC development. As a result, although the HT segments correspond to MAC-destined regions in the target genomes and are therefore very gene-rich, the genes they contain are unlikely to be expressed in *P. sonneborni*. In total, among the 521 HT segments, we identified only 35 genes located in the MAC-destined compartment of *P. sonneborni* (Suppl. Table 1). Comparison with their homologs in other *aurelia* species showed that at least 25 of them are pseudogenes (containing nonsense or frameshift mutations), and among the ten that appear to be intact, only four show evidence of transcription (> 1RPKM) in vegetative cells. Thus, even though HT segments detected in *P. sonneborni* included several thousand genes, most of them have lost their functionality. This implies that *P. sonneborni* was the recipient, and not the donor of the horizontal transfers.

We did not detect any case of a gene from the donor genome integrated by recombination at its homologous locus in the *P. sonneborni* MIC genome. Most of the HT segments are located in the MIC-limited compartment, and hence away from genic regions, and the few (pseudo)genes that we detected in the MAC-destined compartment are most often located towards the end of MAC chromosomes (Suppl. Table 1). Thus, these HT events did not lead to the replacement of recipient genes by their donor homologues, but to the addition of new gene copies (i.e. gene duplication via HT).

P. sonneborni has received HT from four distinct donor lineages (Figure 1A). Several observations indicate that the detected HT segments result from numerous independent genetic transfers, and not from just four occasional events. Indeed, the level of sequence divergence observed between *P. sonneborni* HT segments and the target genomes varies widely among HT segments, ranging from ~0% to ~12%, even among segments coming from the same lineage (Figure 2). Two non-exclusive hypotheses can be proposed to explain this large variance: i) some HT events are more ancient than others, and hence have had more time to

accumulate divergence between the donor and the recipient lineages, or ii) HT segments have about the same age, but originated from multiple related lineages that vary in their level of divergence from the reference species used in our analyses. One noteworthy observation is that the length of HT segments tends to decrease as the sequence divergence increases (Figure 2): for a given donor lineage, the longest HT segments (approaching the size of full-length chromosomes) are generally less divergent from the reference target than other HT segments, whereas the most divergent HT segments are always relatively short. The most parsimonious explanation for this pattern is that after the transfer of full-length chromosomes, HT segments progressively shorten via chromosomal rearrangements (inversions, translocations, deletions) and concomitantly diverge via point substitutions. This therefore suggests that the variance in level of divergence mainly reflects the variability in the age of HT events. In any case, these observations imply that *P. sonneborni* recurrently acquired DNA coming from distant lineages.

In principle, when two species are able to form hybrids, introgressions may occur in both directions. Unfortunately, we could not assess the presence of HT segments related to *P. sonneborni* in any of the four lineages from which it received DNA (represented in our data set by *P. tredecaurelia, P. quadecaurelia, P. novaurelia* and *P. sexaurelia*): for the first three, we do not have a MIC genome assembly, and for *P. sexaurelia*, we do not have an appropriate sister group to detect HT from *P. sonneborni*. Thus, the question of whether transfers occur in both directions remains to be investigated.

DISCUSSION

Recurrent hybridization with genetically highly divergent species.

Hybridization between closely related species is quite common in nature, but often leads to inviable or sterile offspring (Edelman and Mallet 2021). Different types of processes cause this decreased fitness of hybrids. First, after their divergence, sister species tend to accumulate genetic incompatibilities (i.e. alleles that are neutral or beneficial in one species but that interact negatively with the alleles in the other species), so-called Bateson–Dobzhansky–Muller incompatibilities (BDMIs) (Orr 1996). Notably, as each species has adapted to its specific ecological niche, the mixed set of alleles in the hybrids may turn out to be unfit in both environments. Second, chromosomal rearrangements and sequence divergence impair homologous recombination, and thereby cause incorrect chromosome segregation during meiosis in hybrids (Bozdag and Ono 2022). Hence, the probability of forming fertile hybrids (and the potential of inter-species gene flow) decreases with increasing genetic distance (Edelman and Mallet 2021). Typically, in animals and plants, the amount of gene flow across populations/species drops drastically beyond 2% of net synonymous divergence (Roux et al. 2016; Monnet et al. 2025) and in budding yeast, hybrid viability falls to almost zero above 10% of genetic divergence (Bozdag and Ono 2022).

Here we show that a large fraction of the *P. sonneborni* genome has been acquired via recurrent events of horizontal transfer (HT) from multiple lineages. Cases of horizontal transfers had already been reported in the *P. aurelia* complex (Sellis et al. 2021), but they were of relatively short size (<3.5kb) and corresponded to transposable elements. Such transfers might be mediated by vectors (e.g. viruses), and therefore do not necessarily involve inter-specific mating. The size of the longest HT segments detected in *P. sonneborni* (> 200 kb), is in the range of germline chromosome sizes (300 kb to 1 Mb) (Arnaiz et al. 2025). This implies that

they were most probably transferred as full-length chromosomes, through inter-specific conjugation. Thus, our observations reveal that the *P. sonneborni* lineage recurrently mated and produced fertile offspring with very distant species (synonymous divergence: dS~0.8 to 1 substitution per site). For comparison, this corresponds to the genetic distance between the most divergent mammalian taxa (e.g. monotremes vs. placental mammals (Warren et al. 2008)).

Known examples of distant species able to produce viable hybrid offspring are quite rare. The three most extreme cases reported so far are two species of fern that diverged ~60 Mya (Rothfels et al. 2015), sturgeon and paddlefish that diverged ~184 Mya (Káldy et al. 2020), and two gar fish species that diverged ~100 Mya (Brownstein et al. 2024). The latter case is particularly striking, because unlike the other two, the hybrids are not only viable, but also fertile (Brownstein et al. 2024). It has been suggested that the absence of strong BDMI in gars results from the fact that their genomes evolve extremely slowly (Brownstein et al. 2024). The same explanation probably holds for the two other cases: the median dS between sturgeon and paddlefish is only 0.12 (Redmond et al. 2023), and in the fern, the dS between the two parental genomes is 0.1 (Rothfels et al. 2015). Thus, even though these species have been separated for a very long period, they are genetically quite close. Although introgressions are rare at this level of genetic divergence (Roux et al. 2016; Monnet et al. 2025), several cases have been reported. For instance, gene flow has been observed in the *Drosophila yakuba* clade (~3 Mya of separation), or in *Ciona* (sea squirts; 1.5-2 Mya of separation), between species showing a dS of 0.12 to 0.14 (Turissini and Matute 2017; Fraïsse et al. 2022). But to our knowledge, *P. sonneborni* is the first reported case of a eukaryote showing promiscuous sexual behavior with genetically highly diverged species (dS~1).

Remaining the same despite hybridization with distant species.

Our observations raise two questions: how did these hybrids escape the negative consequences of BDMIs and how were they able to complete meiosis? The answer to the first point probably lies in the particular nuclear dimorphism of ciliates. Indeed, the HT segments that we identified in the genome of *P. sonneborni* are essentially confined to its MIC-limited compartment and hence are not expressed, which effectively hides genetic incompatibilities. In *P. aurelia* species, the distinction between segments that are retained or eliminated during MAC development relies on a comparison between the maternal MIC and the maternal MAC, via small RNAs that are produced during meiosis and are still present in the cytoplasm of the zygote after conjugation. The so-called scnRNAs mediate a genomic subtraction that targets the elimination from the new developing MAC of all sequences that are present in the maternal MIC but absent from the maternal MAC (Bétermier et al. 2023; Balan et al. 2024) (Figure 3). We conjecture that in *P. sonneborni*, a similar mechanism also leads to the elimination of sequences that are present in the paternal MIC but absent from the maternal MAC. This would explain how, when a conjugation event occurs between *P. sonneborni* and a distantly related species, the chromosomes from the other species are systematically eliminated from the developing MAC in the F1 hybrid carrying the *P. sonneborni* cytoplasm. In other words, the promiscuous behavior observed in *P. sonneborni* may have been facilitated by the capacity of *P. aurelia* species to exclude non-self DNA from the zygotic MAC, based on the content of the maternal MAC. Interestingly, a similar situation has been reported in a dipteran insect, which possesses two chromosomes that have been acquired by introgression from a distant lineage, but are restricted to the germline, and hence are not expressed in the soma (Hodson et al. 2022). To answer the second question, we propose that these viable hybridization events may have resulted from the formation of allotetraploid F1 hybrids. Indeed, when a whole genome

duplication (WGD) occurs by hybridization between distant species, the two sets of pairs of homologs can be readily distinguished by the meiotic recombination machinery, thereby allowing the proper segregation of chromosomes during meiosis. Consequently, even though the *P. sonneborni* lineage appears to have a promiscuous behavior, the descendants of inter-specific mating do not display a hybrid phenotype, because i) the acquired DNA gets immediately eliminated from the MAC and hence is not expressed, and ii) the introgressed chromosomes are divergent enough so as not to interfere with the segregation of their own chromosomes during meiosis in the MIC.

Given that the pattern of DNA elimination during MAC development is maternally inherited, one prediction of our model is that the mitochondria of the donor species should not be introgressed in the recipient species (Figure 3). Consistent with this prediction, the position of *P. sonneborni* in the phylogeny of *P. aurelia* mitochondrial genomes (Arnaiz et al. 2020) is in total agreement with that inferred from the MAC nuclear genome (Sawka-Gądek et al. 2021). This indicates a perfectly vertical transmission of mitochondria in the ancestry of *P. sonneborni*, despite recurrent hybridization events.

Although our model explains how paramecia might escape the deleterious consequences of inter-specific mating, it leaves several open questions. First, it does not provide any clue regarding the *raison d'être* of these recurrent introgressions. HT segments are mostly confined to the MIC-limited compartment, and hence cannot be expressed, and the few introgressed genes found in the MAC-destined compartment have almost all been pseudogenized. Thus, introgressions appear to make very little contribution to the adaptive potential of *P. sonneborni*. Second, this model predicts that inter-specific mating can be tolerated only if it involves a species that is genetically distant enough to be recognized as non-self during MAC development and meiosis (Figure 3). Hence, selection should still favor the evolution of traits that prevent inter-specific pairing. And in fact, in lab conditions, *P. sonneborni* does not conjugate with any of the other species of the *P. aurelia* complex, not even with the closely related *P. jenningsi* species (Aufderheide et al. 1983; Przyboś et al. 2014). So, what does make these conjugation events possible in nature? Little is known regarding the ecology of *P. sonneborni*. Although this species is rare, it is distributed across very wide geographical areas (the three strains reported so far are from Texas, Dominican republic and Cyprus) (Aufderheide et al. 1983; Przyboś et al. 2014; Tarcz et al. 2023), with limited genetic diversity (0.8% synonymous site polymorphism between the Texas and Cyprus strains (Przyboś et al. 2015)). The high frequency of introgression coming from *P. sexaurelia, P. tredecaurelia*, and *P. quadecaurelia* lineages might be due to the fact that like *P. sonneborni*, these species are preferentially found in warm climate zones (Melekhin et al. 2024). However, *P. octaurelia* is also common in tropical environments and was found in sympatry with *P. sonneborni* (Tarcz et al. 2023), and yet we did not detect any HT from this lineage (Figure 1A).

A third intriguing point is the fact that this promiscuous behavior is observed only in the *P. sonneborni* lineage. For *P. sonneborni* and *P. sexaurelia*, our experimental design provided the same power to detect HT events originating from subclade A. Yet, whereas we identified ~6,000 kb of HT segments coming from subclade A in *P. sonneborni,* we detected only 76 kb in *P. sexaurelia* (Fig. 1A,C). Moreover, we detected very few HT segments in the five other MIC genomes analyzed, and they are all short, suggesting that they may result from a different mode of horizontal transfer (e.g. via viral vectors). Thus, *P. sonneborni* appears to be unique among paramecia in showing such a high rate of introgression in its ancestry.

We speculate that the large size of the MIC-limited compartment in *P. sonneborni* (135 Mb compared to ~30 Mb in other *aurelia* species) might result from its propensity to hybridize with distant species. Although the HT segments that we identified in *P. sonneborni* cover only 7.4% of its MIC-limited compartment, they probably represent only the visible tip of the

iceberg. Indeed, our study design allowed the detection of HT segments only if they encompassed MAC-destined regions from the donor genome and if they were closely related (>80% sequence identity) to the available reference genomes. Given that sequences in the MIC-limited compartment cannot be expressed, they diverge rapidly, like pseudogenes. It is therefore plausible that a large fraction of the MIC-limited compartment of *P. sonneborni* corresponds to ancient HT segments that are no longer recognizable. Thus, the large size of the genome of *P. sonneborni* might simply reflect the recurrent gains of full sets of chromosomes via allotetraploidization, followed by progressive shrinking through deletions or chromosome loss.

The genome of *P. sonneborni* provides an interesting perspective on the definition of species. The model we propose (Figure 3) explains how it is possible for a species to persist, despite recurrent hybridization with very distant lineages. Indeed, owing to nuclear dimorphism, hybrids remain phenotypically identical to their maternal lineage, even if they have received a full set of chromosomes from a very divergent species. Moreover, they also remain interfertile with their parental lineage. Thus, as long as gene flow is maintained in the MAC-destined compartment (which represents only ~35% of the *P. sonneborni* genome), individuals can be considered as being part of the same species, even if their MIC-limited compartment is not shared. This tentative extension of the definition of a species echoes Sonneborn's own hesitation in recognizing that what he initially called "varieties" of *P. aurelia* (Sonneborn 1937) are actually sexually incompatible species (Sonneborn 1975). Ironically, the promiscuous sibling is the one that bears the name of the father of *Paramecium* genetics.

MATERIALS AND METHODS

Detection of horizontal transfer across *aurelia* species

We analyzed the MIC genome of each species *(P. tetraurelia, P. octaurelia, P. biaurelia, P. pentaurelia, P. primaurelia, P. sexaurelia* and *P. sonneborni*) (Arnaiz et al. 2025) to search for segments that have been subject to horizontal transfer (HT) across the *aurelia* complex (see Suppl. Table 2 for a full list of strains and genome assemblies used in this study). For this, the MIC genome of each focal species was cut into 500 bp-long windows, that were compared to all genomes (MAC and MIC, if available) of the 12 other aurelia species with BLASTN (version 2.5.0+; (Camacho et al. 2009)) using the following parameters:

```
blastn -query <infile> -db <db> -out <File_Out> -task blastn -evalue 1e-6 -
max_target_seqs 1
```

For each window, we retained the best blast hit on each of the target genomes. We then selected windows having their best hit with a species that is not a sister-group of the focal species (see phylogenetic tree in Figure 1A). We retained windows with a high level of sequence similarity (>80% identity), significantly higher than with any other species (> mean + 2 standard deviations). These stringent criteria were chosen to identify windows with strong evidence of recent HT.
Then, starting from these 'seed windows', we extended the HT segment by aggregating neighbor windows (< 5kb apart) having their best hit on the same scaffold of the non-sister species. The extension was stopped as soon as a seed window matching another scaffold was reached or when the distance to the next window matching the same scaffold exceeded 5 kb. Finally, we retained HT segments that contained at least 4 seed windows, and that included at

least 70% of windows having their best hits on the same scaffold. The threshold of 4 seed windows was chosen to ensure that HT segments are at least 2kb-long.

We first performed the HT search by comparing MIC windows of the focal species (query) against the MAC genomes of the 12 other *aurelia* species (targets). We then repeated the analyses using as targets the MIC instead of the MAC genomes, for the 7 species for which both the MAC and the MIC were available. In *P. sonneborni*, this second strategy identified 275 additional candidate HT segments, all matching with the MIC of *P. sexaurelia*. However, given that we only have the MAC but not the MIC genome of *P. jenningsi*, we could not ascertain that these segments were absent from the sister-group. In the other species, both strategies gave very similar results (with at most 8 new HT segments identified). To be conservative, the analyses presented in the main text were restricted to HT segments detected with the first strategy (MAC targets).
The median synonymous divergence (dS) between *P. jenningsi* and *P. sonneborni* was computed on the set of 473 one-to-one orthologs reported in a previous study (Sawka-Gądek et al. 2021).

Analysis of HT genes located in the MAC-destined compartment

Based on annotations of the *P. sonneborni* MAC genome assembly (Gout et al. 2023), we identified 39 gene predictions located in HT segments encompassing MAC-destined regions. For each of them, we used BLASTP to identify homologs in other species of the *P. aurelia* complex. We computed multiple alignments to investigate the conservation of each protein family. In 36 out of 39 cases, we identified nonsense and/or frameshift mutations in the *P. sonneborni* predicted gene with respect to its homologs. We also analyzed the expression level of these predicted genes using RNAseq data from *P. sonneborni* vegetative cell cultures (Sawka-Gądek et al. 2021).

Data availability

All genome and annotation files are available from the ParameciumDB download section (https://paramecium.i2bc.paris-saclay.fr/download/). The complete list of detected HT segments and R scripts required to reproduce the figures are available on Zenodo (https://doi.org/10.5281/zenodo.18684075).

ACKNOWLEDGEMENTS

This work was performed using the computing facilities of the CC LBBE/PRABI. We thank Carina Mugal and Mark Stoneking for helpful comments on the manuscript. This work was supported by the Centre National de la Recherche Scientifique, and by the Agence Nationale de la Recherche (ANR-18-CE12-0005/LaMarque). Work in SD laboratory is supported by the Agence Nationale de la Recherche (ANR-23-CE12-0027; ANR-25-CE12-7757), and by the Fondation de la Recherche Medicale (https://www.frm.org)(Equipe FRM EQU202203014643).

REFERENCES

Arnaiz O, Guérin F, Couloux A, Miró-Pina C, Pellerin G, Nekrasova I, Amselem J, Aury J-M, Bhullar S, Frapporti A, et al. 2025. The tiny germline chromosomes of

Paramecium aurelia have an exceptionally high recombination rate and are capped by a new class of Helitrons. 2025.11.06.686955. https://www.biorxiv.org/content/10.1101/2025.11.06.686955v1 (Accessed February 3, 2026).

Arnaiz O, Meyer E, Sperling L. 2020. ParameciumDB 2019: integrating genomic data across the genus for functional and evolutionary biology. *Nucleic Acids Res* **48**: D599–D605. https://doi.org/10.1093/nar/gkz948 (Accessed August 14, 2025).

Aufderheide KJ, Daggett P-M, Nerad TA. 1983. Paramecium sonneborni n. sp., a New Member of the Paramecium aurelia Species-Complex1. *J Protozool* **30**: 128–131. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1550-7408.1983.tb01046.x (Accessed April 23, 2024).

Aury J-M, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Ségurens B, Daubin V, Anthouard V, Aiach N, et al. 2006. Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. *Nature* **444**: 171–178. http://www.nature.com/articles/nature05230 (Accessed March 27, 2018).

Balan T, Lerner LK, Holoch D, Duharcourt S. 2024. Small-RNA-guided histone modifications and somatic genome elimination in ciliates. *WIREs RNA* **15**: e1848. https://onlinelibrary.wiley.com/doi/abs/10.1002/wrna.1848 (Accessed February 2, 2026).

Beisson J, Betermier M, Bre M-H, Cohen J, Duharcourt S, Duret L, Kung C, Malinsky S, Meyer E, Preer JR, et al. 2010. Paramecium tetraurelia: The Renaissance of an Early Unicellular Model. *Cold Spring Harb Protoc* **2010**: pdb.emo140-pdb.emo140. http://www.cshprotocols.org/cgi/doi/10.1101/pdb.emo140 (Accessed March 27, 2018).

Bétermier M, Klobutcher LA, Orias E. 2023. Programmed chromosome fragmentation in ciliated protozoa: multiple means to chromosome ends. *Microbiol Mol Biol Rev* **0**: e00184-22. https://journals.asm.org/doi/10.1128/mmbr.00184-22 (Accessed November 29, 2023).

Bozdag GO, Ono J. 2022. Evolution and molecular bases of reproductive isolation. *Curr Opin Genet Dev* **76**: 101952.

Brownstein CD, MacGuigan DJ, Kim D, Orr O, Yang L, David SR, Kreiser B, Near TJ. 2024. The genomic signatures of evolutionary stasis. *Evolution* **78**: 821–834. https://doi.org/10.1093/evolut/qpae028 (Accessed August 13, 2025).

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

De Souza BA, Dias RJP, Senra MVX. 2020. Intrageneric evolutionary timing and hidden genetic diversity of Paramecium lineages (Ciliophora: Oligohymenophorea). *Syst Biodivers* **18**: 662–674. https://doi.org/10.1080/14772000.2020.1769225 (Accessed July 30, 2025).

Edelman NB, Mallet J. 2021. Prevalence and Adaptive Impact of Introgression. *Annu Rev Genet* **55**: 265–283.

Fokin SI, Przyboś E, Chivilev SM, Beier CL, Horn M, Skotarczak B, Wodecka B, Fujishima M. 2004. Morphological and molecular investigations of Paramecium schewiakoffi sp. nov. (Ciliophora, Oligohymenophorea) and current status of distribution and taxonomy of Paramecium spp. *Eur J Protistol* **40**: 225–243. http://linkinghub.elsevier.com/retrieve/pii/S0932473904000306 (Accessed March 27, 2018).

Fraïsse C, Le Moan A, Roux C, Dubois G, Daguin-Thiebaut C, Gagnaire P-A, Viard F, Bierne N. 2022. Introgression between highly divergent sea squirt genomes: an adaptive breakthrough? *Peer Community J* **2**. https://peercommunityjournal.org/articles/10.24072/pcjournal.172/ (Accessed August 13, 2025).

Gout J-F, Hao Y, Johri P, Arnaiz O, Doak TG, Bhullar S, Couloux A, Guérin F, Malinsky S, Potekhin A, et al. 2023. Dynamics of Gene Loss following Ancient Whole-Genome Duplication in the Cryptic Paramecium Complex. *Mol Biol Evol* **40**: msad107. https://doi.org/10.1093/molbev/msad107 (Accessed April 18, 2024).

Hodson CN, Jaron KS, Gerbi S, Ross L. 2022. Gene-rich germline-restricted chromosomes in black-winged fungus gnats evolved through hybridization. *PLOS Biol* **20**: e3001559. https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3001559 (Accessed November 12, 2024).

Johri P, Krenek S, Marinov GK, Doak TG, Berendonk TU, Lynch M. 2017. Population Genomics of Paramecium Species. *Mol Biol Evol* **34**: 1194–1216. https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msx074 (Accessed March 27, 2018).

Káldy J, Mozsár A, Fazekas G, Farkas M, Fazekas DL, Fazekas GL, Goda K, Gyöngy Z, Kovács B, Semmens K, et al. 2020. Hybridization of Russian Sturgeon (Acipenser gueldenstaedtii, Brandt and Ratzeberg, 1833) and American Paddlefish (Polyodon spathula, Walbaum 1792) and Evaluation of Their Progeny. *Genes* **11**: 753. https://www.mdpi.com/2073-4425/11/7/753 (Accessed August 12, 2025).

Long H, Johri P, Gout J-F, Ni J, Hao Y, Licknack T, Wang Y, Pan J, Jiménez-Marín B, Lynch M. 2023. Paramecium Genetics, Genomics, and Evolution. *Annu Rev Genet* **57**: 391–410.

McGrath CL, Gout J-F, Johri P, Doak TG, Lynch M. 2014. Differential retention and divergent resolution of duplicate genes following whole-genome duplication. *Genome Res* **24**: 1665–1675. http://genome.cshlp.org/lookup/doi/10.1101/gr.173740.114 (Accessed March 27, 2018).

Melekhin M, Potekhin A, Gentekaki E, Chantangsi C. 2024. Paramecium (Oligohymenophorea, Ciliophora) diversity in Thailand sheds light on the genus biogeography and reveals new phylogenetic lineages. *J Eukaryot Microbiol* **71**: e13004. https://onlinelibrary.wiley.com/doi/abs/10.1111/jeu.13004 (Accessed July 30, 2025).

Monnet F, Postel Z, Touzet P, Fraïsse C, Van de Peer Y, Vekemans X, Roux C. 2025. Rapid establishment of species barriers in plants compared with that in animals. *Science*

389: 1147–1150. https://www.science.org/doi/full/10.1126/science.adl2356 (Accessed February 3, 2026).

Orr HA. 1996. Dobzhansky, Bateson, and the Genetics of Speciation. *Genetics* **144**: 1331–1335.

Potekhin A, Mayén-Estrada R. 2020. Paramecium Diversity and a New Member of the Paramecium aurelia Species Complex Described from Mexico. *Diversity* **12**: 197. https://www.mdpi.com/1424-2818/12/5/197 (Accessed July 30, 2025).

PrzyboŚ E, Tarcz S. 2019. Global molecular variation of Paramecium jenningsi complex (Ciliophora, Protista): a starting point for further, detailed biogeography surveys. *Syst Biodivers* **17**: 527–539. https://doi.org/10.1080/14772000.2019.1643424 (Accessed July 30, 2025).

Przyboś E, Tarcz S, Rautian M, Lebedeva N. 2014. The first European stand of *Paramecium sonneborni* (*P. aurelia* complex), a species known only from North America (Texas, USA). *Eur J Protistol* **50**: 236–247. https://www.sciencedirect.com/science/article/pii/S0932473914000182 (Accessed July 30, 2025).

Przyboś E, Tarcz S, Rautian M, Sawka N. 2015. Delimiting Species Boundaries within a Paraphyletic Species Complex: Insights from Morphological, Genetic, and Molecular Data on Paramecium sonneborni (Paramecium aurelia species complex, Ciliophora, Protozoa). *Protist* **166**: 438–456. http://linkinghub.elsevier.com/retrieve/pii/S1434461015000449 (Accessed March 27, 2018).

Redmond AK, Casey D, Gundappa MK, Macqueen DJ, McLysaght A. 2023. Independent rediploidization masks shared whole genome duplication in the sturgeon-paddlefish ancestor. *Nat Commun* **14**: 2879. https://www.nature.com/articles/s41467-023-38714-z (Accessed August 13, 2025).

Rothfels CJ, Johnson AK, Hovenkamp PH, Swofford DL, Roskam HC, Fraser-Jenkins CR, Windham MD, Pryer KM. 2015. Natural Hybridization between Genera That Diverged from Each Other Approximately 60 Million Years Ago. *Am Nat* **185**: 433–442. https://www.journals.uchicago.edu/doi/10.1086/679662 (Accessed August 12, 2025).

Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. 2016. Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence ed. C. Moritz. *PLOS Biol* **14**: e2000234. http://dx.plos.org/10.1371/journal.pbio.2000234 (Accessed August 8, 2019).

Sawka-Gądek N, Potekhin A, Singh DP, Grevtseva I, Arnaiz O, Penel S, Sperling L, Tarcz S, Duret L, Nekrasova I, et al. 2021. Evolutionary Plasticity of Mating-Type Determination Mechanisms in Paramecium aurelia Sibling Species. *Genome Biol Evol* **13**. https://doi.org/10.1093/gbe/evaa258 (Accessed March 2, 2021).

Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH. 2006. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* **440**: 341–345. https://www.nature.com/articles/nature04562 (Accessed August 6, 2025).

Sellis D, Guérin F, Arnaiz O, Pett W, Lerat E, Boggetto N, Krenek S, Berendonk T, Couloux A, Aury J-M, et al. 2021. Massive colonization of protein-coding exons by selfish genetic elements in Paramecium germline genomes. *PLOS Biol* **19**: e3001309. https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3001309 (Accessed September 30, 2021).

Sonneborn T. 1975. The Paramecium aurelia complex of 14 sibling species. *Trans Am Microsc Soc* **94**: 155–178.

Sonneborn TM. 1937. Sex, Sex Inheritance and Sex Determination in Paramecium Aurelia. *Proc Natl Acad Sci U S A* **23**: 378–385.

Stankowski S, Ravinet M. 2021. Defining the speciation continuum. *Evolution* **75**: 1256–1273. https://doi.org/10.1111/evo.14215 (Accessed August 6, 2025).

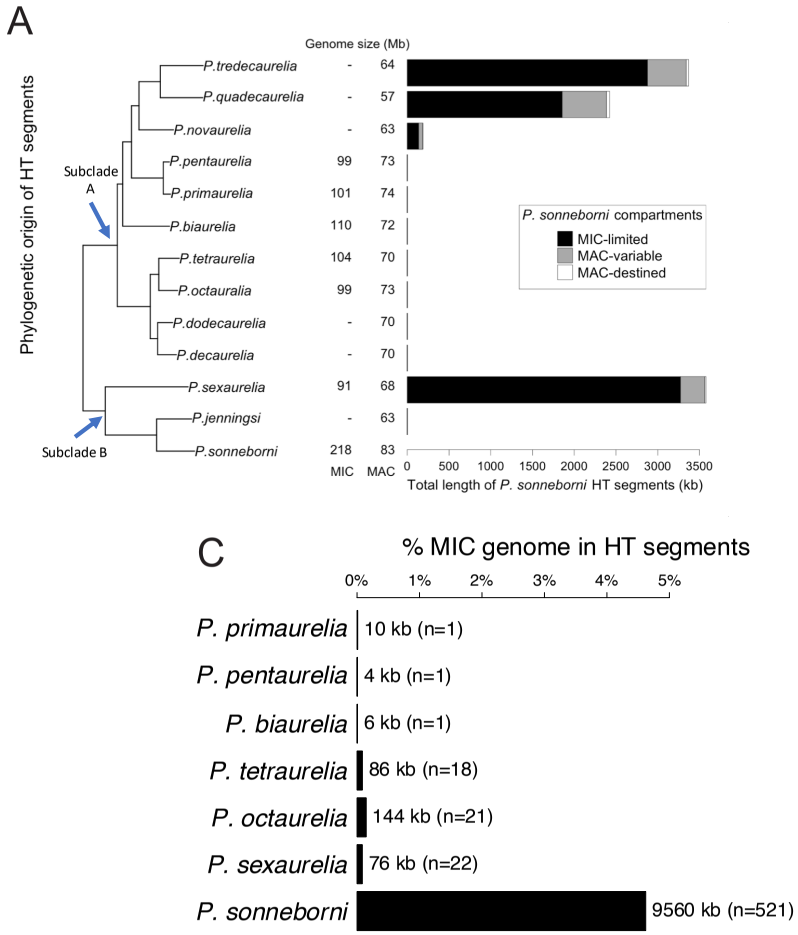
Tarcz S, Surmacz M, Przyboś E. 2023. Sampling hidden microbial eukaryotic biodiversity in the tropics: new insights from the Paramecium aurelia complex (Ciliophora, Protozoa). *Folia Biol (Praha)* **71**: 159–170.

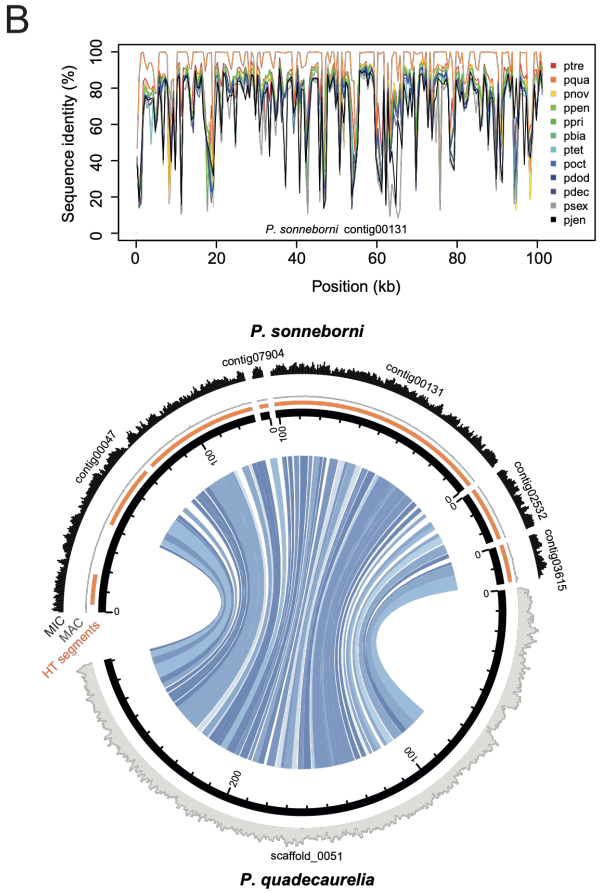
Turissini DA, Matute DR. 2017. Fine scale mapping of genomic introgressions within the Drosophila yakuba clade. *PLOS Genet* **13**: e1006971. https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006971 (Accessed August 13, 2025).

Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, Grützner F, Belov K, Miller W, Clarke L, Chinwalla AT, et al. 2008. Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453**: 175–183. http://www.nature.com/doifinder/10.1038/nature06936 (Accessed March 27, 2018).

Zhang X, Gong R, Jiang Y, Lu X, Wu C, Wang L, Ma H, Zhang Z, Song W, Al-Rasheid KAS, et al. 2023. Nuclear events during conjugation in the poorly studied model ciliate *Paramecium jenningsi*. *Water Biol Secur* **2**: 100201. https://www.sciencedirect.com/science/article/pii/S2772735123000872 (Accessed August 14, 2025).

FIGURES


A
Phylogenetic origin of HT segments
Genome size (Mb)
P.tredecaurelia
P.quadecaurelia
P.novaurelia
P.pentaurelia
P.primaurelia
P.biaurelia
P.tetraurelia
P.octauralia
P.dodecaurelia
P.decaurelia
P.sexaurelia
P.jenningsi
P.sonneborni
Subclade A
Subclade B
MIC
MAC
P. sonneborni compartments
MIC-limited
MAC-variable
MAC-destined
Total length of P. sonneborni HT segments (kb)


B
Sequence identity (%)
Position (kb)
P. sonneborni contig00131
ptre
pqua
pnov
ppen
ppri
pbia
ptet
poct
pdod
pdec
psex
pjen
P. sonneborni
contig07904
contig00131
contig00047
contig02532
contig03615
MIC
MAC
HT segments
scaffold_0051
P. quadecaurelia

C
% MIC genome in HT segments
P. primaurelia
10 kb (n=1)
P. pentaurelia
4 kb (n=1)
P. biaurelia
6 kb (n=1)
P. tetraurelia
86 kb (n=18)
P. octaurelia
144 kb (n=21)
P. sexaurelia
76 kb (n=22)
P. sonneborni
9560 kb (n=521)

Figure 1 Horizontal transfer in the *Paramecium aurelia* complex

A. Phylogenetic tree of *Paramecium aurelia* species (modified from (Gout et al. 2023)). For each species, the sizes of the MAC and, if available, MIC genome assemblies are indicated. All the *aurelia* species have MAC genomes around 70 Mb and MIC genomes around 100 Mb in size, with the notable exception of *P. sonneborni* whose MIC genome is more than twice as large (Arnaiz et al. 2025). The barplot indicates the phylogenetic origin of HT segments identified in the *P. sonneborni* genome. HT regions can be localized in three genomic compartments: MIC-limited (black), MAC-variable (dark grey) and MAC-destined (light grey).

B. Example of a region of the *P. sonneborni* MIC genome that presents the hallmarks of HT from other *P. aurelia* species. The upper panel presents the data for *P. sonneborni* contig00131: the contig was cut into 500bp-long non-overlapping windows, and each window was compared (BLASTN) to all other *P. aurelia* MAC genomes. Each colored line indicates the level of sequence identity with their best hit in the corresponding target genome. HT segments were identified in *P. sonneborni* by searching for sets of neighboring windows with conserved synteny, and whose best hits were not with its sister group (*P. jenningsi*, pjen) (see Methods for details). In this example, the entire *P. sonneborni* contig shows a very high level of similarity with the genome of *P. tredecaurelia* (ptre), more than with any other *P. aurelia* species. Species codes are given in Supplemental Table 2.

Lower panel, the circos plot shows the conservation of synteny between five *P. sonneborni* contigs containing HT segments (orange) and the putative donor region of the *P. quadecaurelia* MAC genome (scaffold_0051). Histograms describe DNA-seq coverage for MIC DNA sample (black) and MAC DNA sample (grey). In this example, the HT segments are located in the MIC-limited compartment of *P. sonneborni*: they are well covered in the MIC DNA sample but not in the MAC. The colors of the arcs are defined by the similarity between regions.

C. Percentage of MIC genomes in horizontally transferred (HT) segments. The numbers next to each bar indicate the cumulative size and number of HT segments.

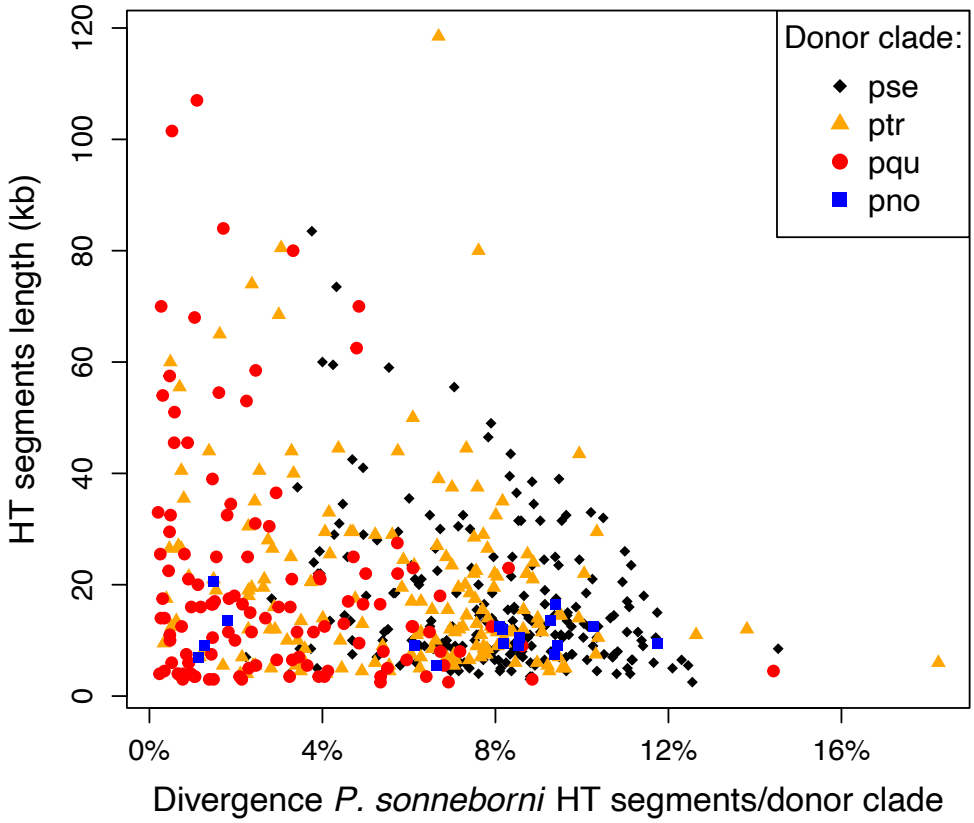


Figure 2 Relationship between the size of HT segments in the *P. sonneborni* genome, and their divergence from the donor lineage.

The origin of the donor clade is indicated by different colors and dot styles. Species codes are given in Supplemental Table 2.

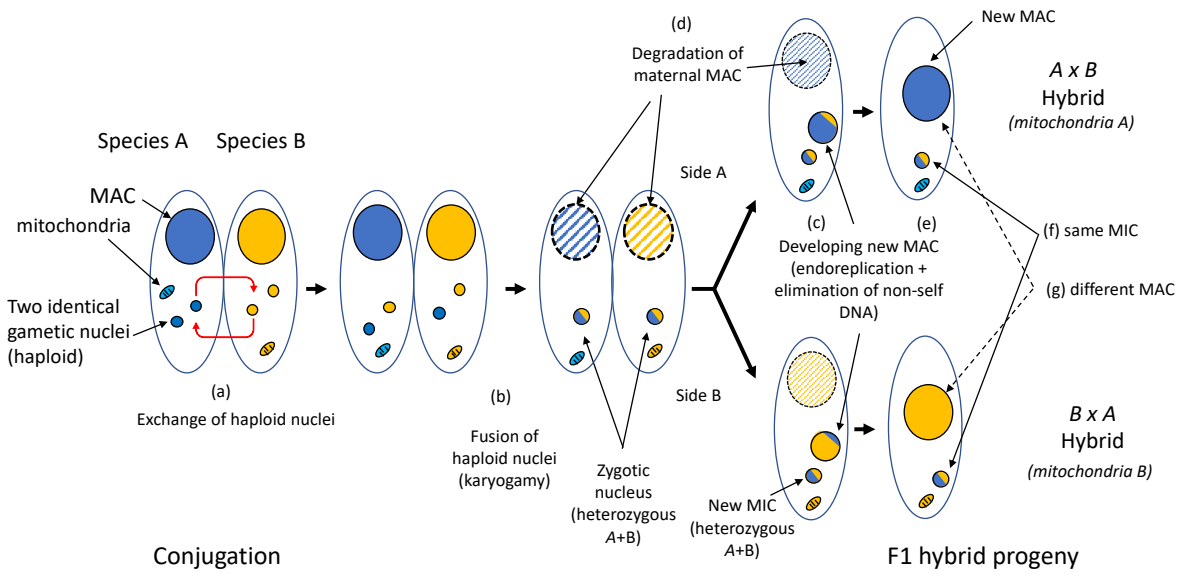


Figure 3 Model: how the nuclear dimorphism in Paramecia might hide interspecific genetic incompatibilities.

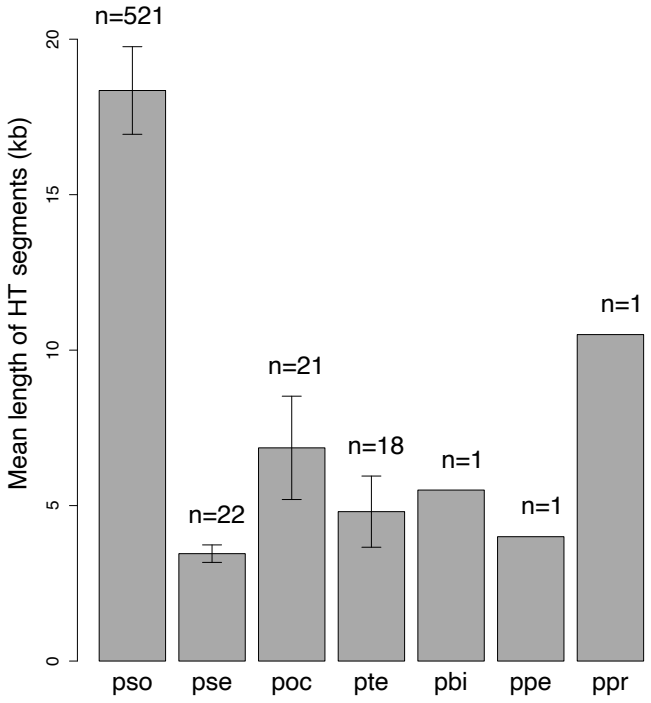
Sexuality in paramecia occurs through conjugation: cells with compatible mating types pair and engage meiosis. Three of the four resulting haploid nuclei are eliminated, and the remaining one undergoes a mitotic division, leading to the formation of two identical haploid gametic nuclei. The two paired cells exchange one haploid nucleus (a), and then, in each cell, the haploid nuclei fuse to form the new zygotic nucleus (b). After two mitotic divisions, one nucleus differentiates to form the new MAC, while the other constitutes the new MIC (c) (for the sake of simplicity, only one MIC nucleus is shown in the figure; see (Zhang et al. 2023) for more details), while the maternal MAC nucleus is progressively degraded (d). During the development of the new MAC, its chromosomes are amplified by endoreplication, and a fraction of the genome (termed MIC-limited) is eliminated. The distinction between segments that are retained or eliminated relies on a comparison between the maternal MIC and the maternal MAC, via small RNAs produced during meiosis: sequences that are present in the maternal MIC but absent from the maternal MAC are eliminated (c) (Balan et al. 2024). Hence, even though the two F1 progeny have exactly the same germline genotype (f), their MAC content may differ, depending on the content of their respective maternal MACs (g). The model that we propose relies on the assumption that during interspecific crosses, sequences that are present in the paternal MIC but absent from the maternal MAC are also eliminated.

This would explain how crosses between distant species can lead to a viable and fertile offspring. Indeed, let us consider a conjugation event between two genetically divergent species (say A and B), and let us focus on the daughter cell from the A side of the cross (a). Its new MIC is allotetraploid, with the full sets of chromosomes from A and B (b). But given that B sequences are divergent from the A sequences present in its maternal MAC, they will all be eliminated during the development of its new MAC (c). Thus, in the side A daughter cell, the new MAC only expresses genes coming from the A genome, thereby effectively hiding any BDMI with the B genome. Hence, although genetically hybrid (A+B), the side A daughter cell should be viable and retain the phenotype of its maternal lineage (A). Moreover, these cells should be able to complete meiosis normally, as the divergence between the two genomes limits the risk of mispairing between homeologous A/B chromosomes. In F2 individuals resulting from a backcross between an AxB hybrid and an A individual, meiosis is expected to proceed normally for A chromosomes, and hence, even though some of their B chromosomes mis-segregate, these F2's should lead to a viable and fertile offspring. This model can account for the introgression of DNA coming from B in the MIC-limited compartment of species A.

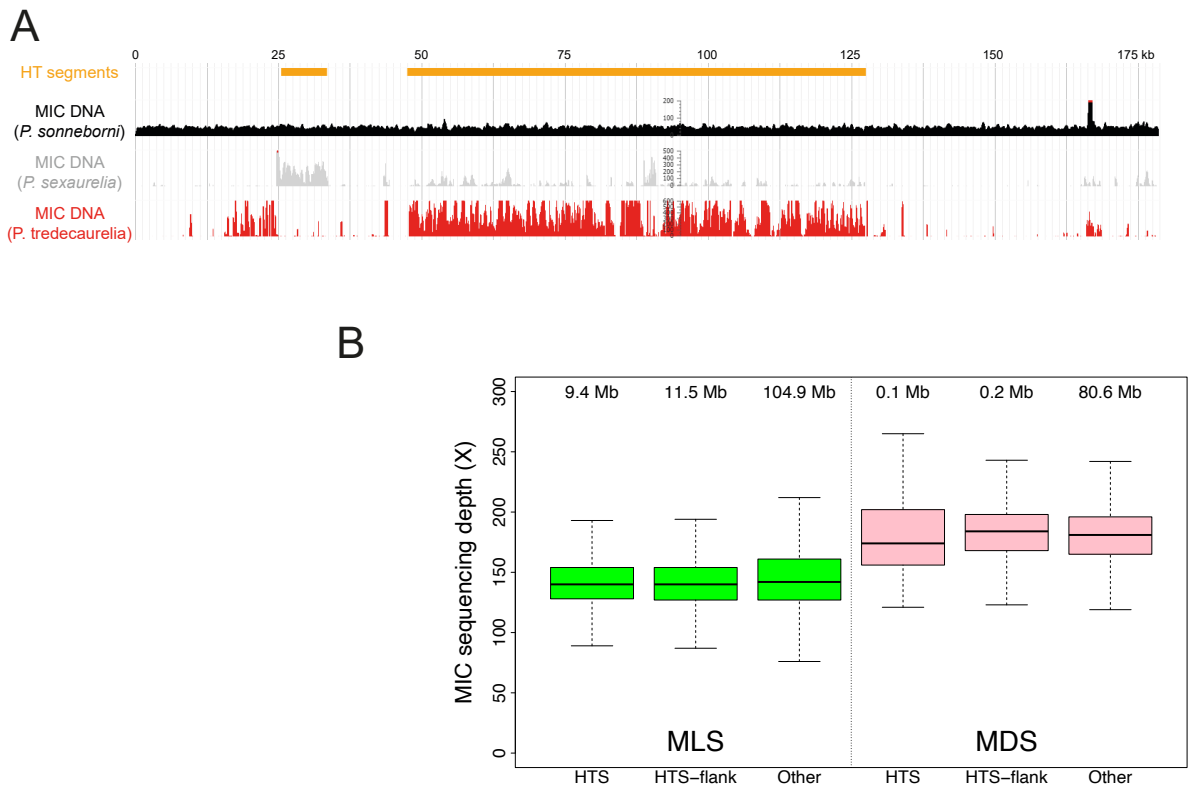
Supplementary Information

Suppl. Table 1: **List of genes identified in HT segments located in the MAC-destined compartment of *P. sonneborni***
Suppl. Table 2: **List of genome assemblies and strains used in this study**

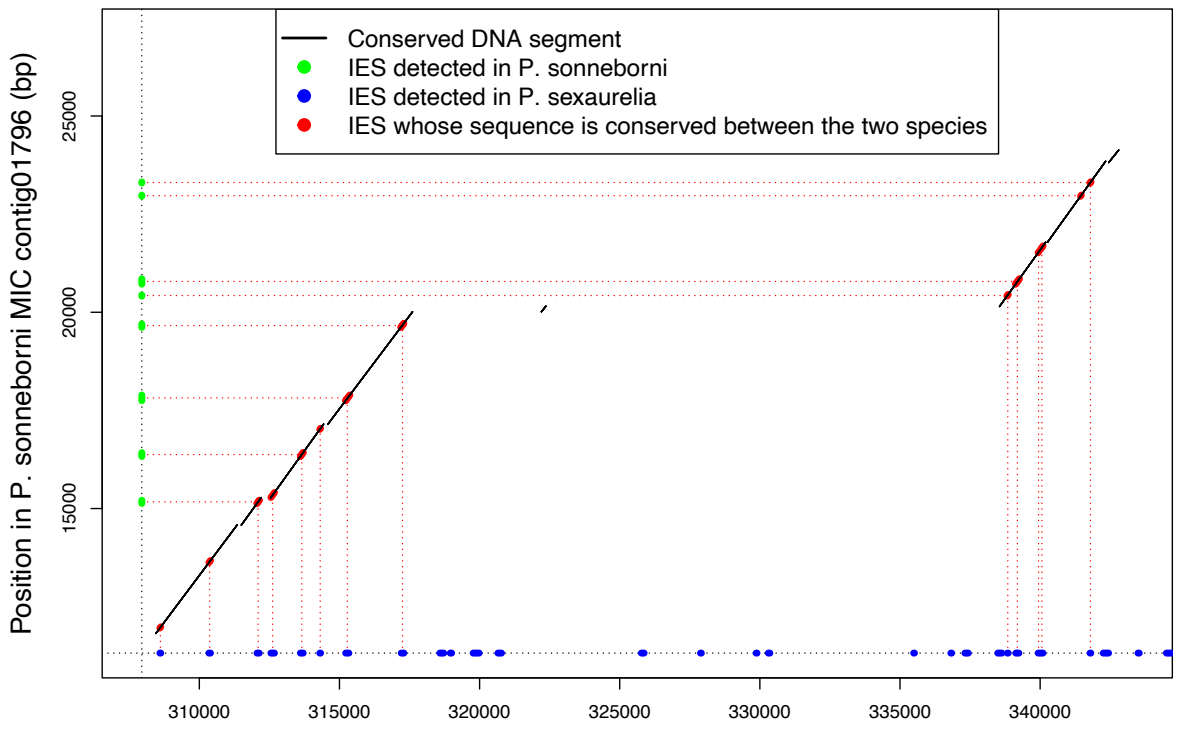
Supplementary Figures



Suppl. Figure 1: **Average length of HT segments detected in the MIC genomes of the 7 analyzed species**. Error bars represent the 95% confidence interval for the mean. Species codes: pso: *P. sonneborni*, pte: *P. tetraurelia*, pbi: *P. biaurelia*, poc: *P. octaurelia*, pse: *P. sexaurelia*, ppr: *P. primaurelia*, ppe: *P. pentaurelia*.



Suppl. Figure 2: **Evidence that HT segments detected in *P. sonneborni* do not result from DNA contamination artefacts**. **A.** Example of a MIC contig of *P. sonneborni* (contig00015) that includes two HT segments (orange rectangles), the first derived from *P. sexaurelia* (scaffold_001.2), the second from *P. tredecaurelia* (scf7180000129386). If HT segments resulted from DNA contamination, this would imply that contig00015 corresponds to an artefactual chimeric assembly of DNA fragments coming from three sources (*P. sonneborni*, *P. sexaurelia* and *P. tredecaurelia*). Hence, the sequencing depths would be expected to differ between the two HT segments (originating from *P. sexaurelia* and *P. tredecaurelia*), and to be lower in HT segments than in flanking regions (*P. sonneborni*). The first profile (black) indicates the sequencing depth for the *P. sonneborni* MIC DNA sample. This profile shows a uniform sequencing depth along the whole contig, including in regions outside of HT segments, in contradiction with the hypothesis of a DNA contamination artefact. As a control, we also mapped sequence reads from MIC DNA samples obtained from *P. sexaurelia* (grey) or *P. tredecaurelia* (red). As expected, we observed a high density of reads from each species mapping on the corresponding HT segment. The heterogeneity of coverage reflects the fact that the sequences of the *P. sexaurelia* and *P. tredecaurelia* strains that we used in this study are not identical to the HT segments present in *P. sonneborni*, again in contradiction with the hypothesis that these HT segments might result from DNA contamination. **B.** Distribution of sequencing depth within and outside HT segments in *P. sonneborni*. The MIC genome assembly was split into 500bp-long non-overlapping windows, which were classified as MIC-limited (MLS) or MAC-destined (MDS), based on their MAC sequencing depth, and then subdivided into three subsets: windows that overlap a HT-segment (HTS), windows that flank a HT-segment (HTS-flank), and windows located on other MIC contigs (i.e. not flanking a HT-segment). Both in MLS and in MDS regions, the distribution MIC sequencing depth (represented by boxplots) is the same within HT-segments as in their flanking regions and the same as in the rest of the genome, consistent with the assumption that HT-segments correspond to authentic *P. sonneborni* MIC DNA. The amount of DNA in each class of windows is given above the corresponding boxplot.



Suppl. Figure 3: **Presence of conserved IESs within HT segments**. This figure presents a dot plot alignment between a *P. sonneborni* MIC contig (contig01796) and the *P. sexaurelia* MIC scaffold (scaffold_103), with which we identified a HT segment (88% sequence identity between the two species). IESs have been previously detected in the MIC genomes of *P. sonneborni* and of *P. sexaurelia* (Sellis et al. 2021). The figure shows that this HT segment encompasses 14 conserved IESs, among which 7 were detected in both species, 6 were detected only in *P. sexaurelia* and 1 only in *P. sonneborni*. Given that IESs are absent from the MAC genome, the presence of conserved IESs within the HT segment indicates that it resulted from the transfer of MIC, rather than MAC DNA.