RESEARCH COMMENTARY

Origin, structure and function of millions of chromosomes present in the macronucleus of unicellular eukaryotic ciliate, *Oxytricha trifallax*: a model organism for transgenerationally programmed genome rearrangements

SUSHIL KUMAR[1]* and RENU KUMARI[2]

[1]*SKA Institution for Research, Education and Development, 4/11 Sarvapriya Vihar, New Delhi 110 016, India*
[2]*National Institute for Plant Genome Research, Aruna Asaf Ali Marg, New Delhi 110 067, India*

[Kumar S. and Kumari R. 2015 Origin, structure and function of millions of chromosomes present in the macronucleus of unicellular eukaryotic ciliate, *Oxytricha trifallax*: a model organism for transgenerationally programmed genome rearrangements. *J. Genet.* **94**, 171–176]

Introduction

The unicellular eukaryotic ciliate protists characteristically contain a germline micronucleus (MIC) and a somatic macronucleus (MAC) in their cytoplasm. The MAC, which is crucial for the pursuit of cellular growth and mitotic divisions, is derived from a postzygotic MIC. The transition from MIC to MAC involves extensive editing of the MIC genome, followed by massive amplification of the residual MAC genome. The spirotrichous ciliate, *Oxytricha trifallax*, has been most extensively deployed as the model system to reveal the mechanism(s) of origin of the architectural complexities of MAC. This has been possible by the application of genome sequencing, synthetic RNA transgenesis and a variety of other genetic techniques. Here, we summarize and discuss the current observations of MAC genome derived from MIC genome, properties of MAC genome structure, function and significance of the novel features described in *O. trifallax*. The differences in the genome organization of *O. trifallax* and its relative alveolate species *Paramecium tetraurelia* and *Plasmodium falciparum* have been described. Aspects of programmed genome rearrangements, MAC genome structure and function requiring further analyses in different ciliate protest have been pointed out.

MIC and MAC genome structures

Figure 1 gives a diagramme of *O. trifallax* cell and the fates of MIC and MAC during asexual and sexual reproduction events. It also shows the origin of MAC from a copy of postzygotic MIC via MAC differentiation process. The MAC is morphologically much larger in size than MIC, although the sizes of haploid MIC and MAC genomes are about 500 and 50 Mb, respectively (Chen *et al.* 2014). MIC diploid genome has about 810 genes that have normal eukaryotic gene structure and are expressed in MIC during conjugal reproduction (Chen *et al.* 2014). Remaining MIC genome specifies transposons, internally eliminated sequences (IESs), satellite sequences and about 18,400 MAC-expressed genes (Swart *et al.* 2013; Chen *et al.* 2014). Among these 18,400 MAC destined genes, about 17,850 are interrupted mostly in their exons by at least one IES; and about 550 genes are IES-less (Chen *et al.* 2014). IES are transposon sequences that have lost their transposases and require in trans the transposase activity(ies) of other types of transposons, coexisting in the genome. IESs disrupt the genes destined for MAC into more than 225,000 segments that have been called as macronucleus destined segments (MDSs). In a gene, the MDSs are not always located contiguously at a locus. Many MDSs are inverted or translocated to distinct gene loci on the same or different chromosome (Bracht *et al*. 2013; Swart *et al*. 2013; Chen *et al*. 2014). The process of differentiation of MIC into MAC involves deletion of IESs, unscrambling and correctly ordering of MDSs specific genewise and loss of transposons. Identical short sequences or pointers (2–20 bp long) of usually 4 bp or 9 bp present at MDS–IES junctions facilitate homologous recombination between pairs and allow correct genewise ordering of MDSs (Bracht *et al*. 2013). This process produces about 16,000 MAC chromosomes (Swart *et al.* 2013). Each chromosome is bound at the end by telomeres, but centromere is absent. Figure 2 diagrammes the process of nanochromosome formation from three contiguous MDSs of a gene (MDS1, MDS2 and MDS3) that are scrambled (-MDS2-MDS3-MDS1-) and one of them

*For correspondence. E-mail: sushil2000_01@yahoo.co.in.

Keywords. genome amplification; internally eliminated sequences; programmed genome editing; polyploidy; RNA-mediated epigenetics; transgenerational inheritance.

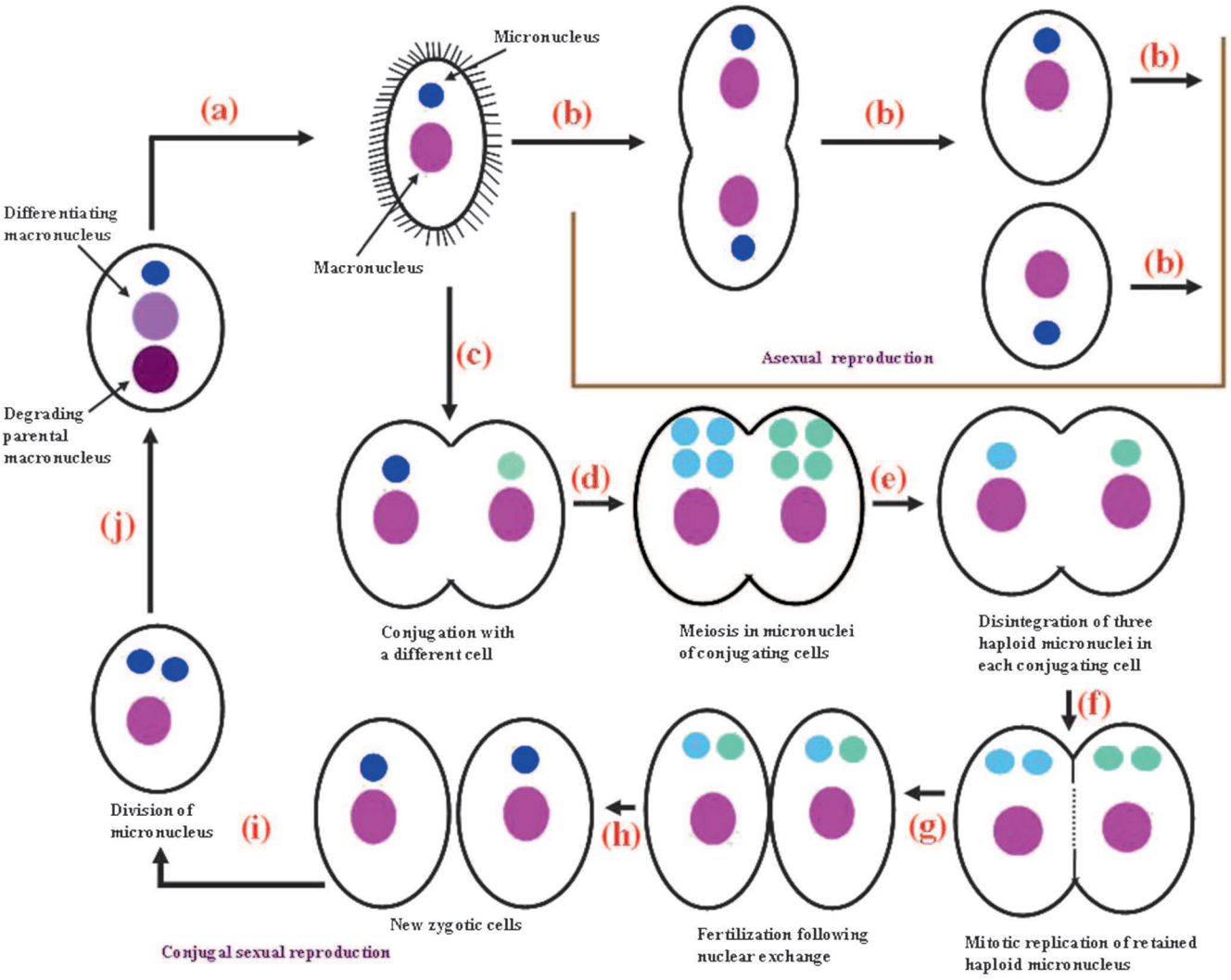

Figure 1. Diagramme of the life cycle of unicellular microbial spirotrichous ciliate protest, *Oxytricha trifallax*. The cells of the organism are large (150 μM in length and 70 μM in width or about 10 times the size of human cell), occur in water bodies and moist soil. Two important characteristics of the species are presence of cilia on surface that aid in movement, attachment and feeding and nuclear dimorphism; a germline diploid MIC that participates in sexual reproduction and a somatic polyploidy MAC that determines asexual cell growth and reproduction (a). The MAC is generated from MIC and the process involves heavy editing of DNA (deletions, inversions and translocations to give rise to thousands of nanochromosomes), followed by their at least thousand-fold amplification (see figures 2 and 3 for the diagrammatic representation of the processes involved in MAC differentiation). Asexual cell divisions occur by fission (b). MIC undergoes mitosis and MAC amitosis (because the chromosomes in MAC lack centromeres). Each of the two daughter cells acquires a MIC and MAC. Sexual reproduction begins when two cells form a cytoplasmic bridge that results in conjugational fusion (c). MIC of each conjugating cell undergoes meiosis to produce four haploid promicronuclei (d). Three promicronuclei disintegrate in each of the conjugating cells (e). The remaining promicronucleus divides mitotically to produce two gametic nuclei, a stationary and a migratory (f). The migratory MIC of a conjugating cells moves into the other cells and vice versa (g). The resident stationary and incoming migratory haploid micronuclei fuse in each conjugating cell to give rise to zygotic nucleus (h). The conjugating cells separate and their micronuclei undergo a round of mitotic division (i). Whereas one MIC becomes the germline MIC and the other develops into MAC (j). At the same time the old (parental) MAC begins to disintegrate while contributing RNA copies of all the nanochromosomes to the differentiating MAC. These serve as resource for piRNAs and RNA scaffolds for guiding genome rearrangement (joining and unscrambling of MAC directed segments (MDSs)). Each cell that underwent conjugation enters into phase of asexual growth and reproduction, regulated by MAC. In *O. trifallax*, sexual reproduction is to generate germline variability and not cell numbers; asexual reproduction is responsible for the population size increase.

is inverted (2SDM). It also shows nanochromosome amplification, leading to polyploidization.

Majority of MAC chromosomes contain only one transcribable and translatable gene (Swart *et al.* 2013) and their average size is 3.7 kb (Swart *et al.* 2013; Chen *et al.* 2014). With some frequency (~10%), certain alternate forms of these small chromosomes arise; these are isoforms that share some chromosomal sequences. Owing to their small size, the

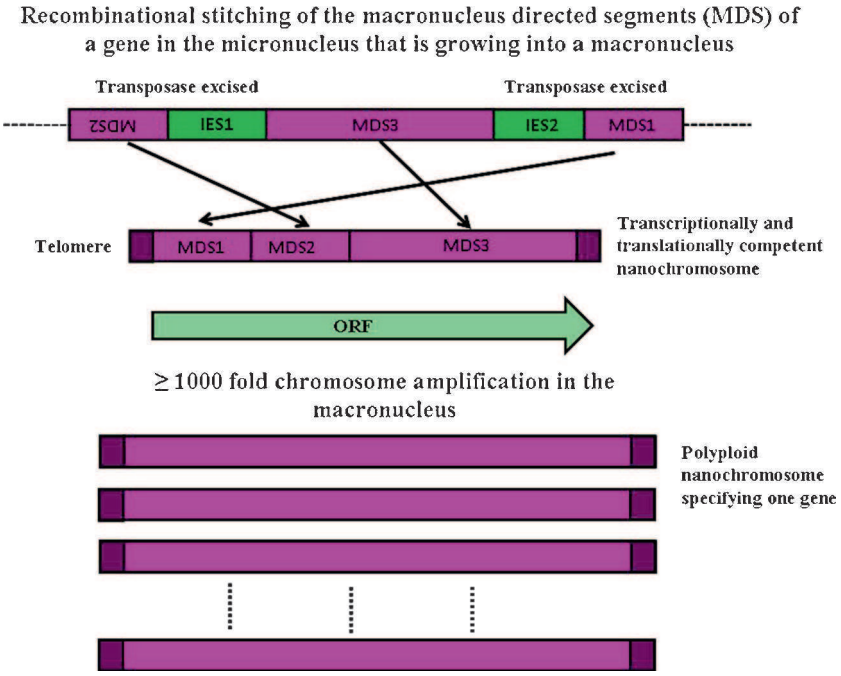

Figure 2. Schematic diagramme of generation of a nanochromosome in the MAC of *O. trifallax*. The MDS as present in the MIC (MDS1–MDS3) are shown in magenta colour. The MDSs are scrambled and one of them (MDS2) is inverted. MDSs are separated by internal eliminated sequences (IESs) that are shown in green colour. In between MDS and IES are present short (2–20 bp) TA-rich pointer direct repeat sequences. In the process of differentiation of MIC into MAC, IES sequences are removed by the action of MIC encoded Tc1/*mariner* transposases and recombination in pointers stiches together MDSs in correct orderly sequences, albeit scrambling and inversion of some of the MDSs giving rise to a transcribable functional gene. Thus single gene nanochromosome formed has telomeres at its ends but lacks centromere. The nanochromosome gets amplified 1000 to 2000 folds. How the nanochromosomes segregate their chromatids during cell divisions in the course of cells undergoing asexual reproduction is not known.

MAC chromosomes have been called the nanochromosomes. Generally, the nanochromosomes that arise from unscrambling of genes are larger in size (4.9 kb). A 22-kb MIC locus is known to be fragmented into 245 MDSs. Unscrambling of these MDSs produce a 13-kb nanochromosome (Chen *et al.* 2014). Formation of nanochromosomes is followed by their 1000–2000 fold amplification giving rise to MAC polyploidy (Bracht *et al.* 2013; Swart *et al.* 2013; Chen *et al.* 2014). Thus, it is estimated from the data reported by Swart *et al.* (2013) that the nanochromosome number in the MAC of *O. trifallax* is about 30 million (15,600 MAC chromosomes $\times 1900$ ploidy level $= 29.64 \times 10^6$ chromosomes).

RNA-guided mechanism of genome rearrangement in differentiating MAC to produce nanochromosomes possessing functional genes

Figure 3 outlines the roles of piRNA and template RNA in precise construction of nanochromosomes (Nowacki *et al.* 2008, 2010, 2011; Nowacki and Landweber 2009; Fang *et al.* 2012; Fang and Landweber 2013; Bracht *et al.* 2013). The paternal MAC serves as the resource of long RNAs, transcripts of entire lengths of both the strands of nanochromosome DNA. These RNAs and also piwi protein, Otiwi1, get transported from the parental MAC to the newly developing MAC. The dsRNAs are processed by dicer-like activity into 27-nt long RNAs. These are called piRNAs, on account of their ability to complex with the piwi protein, Otiwi1. The complex of piRNA $\times$ Otiwi1 marks the MDSs complementary to piRNA. Thereby, the MDSs are shielded against any nuclease attack. On the other hand, long telomere bearing RNA serves as scaffold guide for MDSs to align in proper homology directed order for their recombinational fusion at pointers. Subsequently, IESs get removed to produce a complete gene on a nanochromosome (shown in figure 2). Pointers, the short regions of homology, serve as sites for MDS recombination in the process of IES elimination. Since the length of pointers is small (4 or 9 bp), errors if any (arising from leftover pointer) are corrected by proofreading under template RNA guidance to eventually produce nanochromosome possessing functionally correct gene sequences (Nowacki *et al.* 2008). Once a gene is stitched correctly and incorporated properly on a nanochromosome, the structure of the nanochromosome gets propagated transgenerationally, theoretically over infinite number of generations. A typical nanochromosome comprises of an intron containing coding sequence bounded on the two sides by a small untranslated region followed by a small untranscribed region which is followed by telomere.

Compelling evidence in favour of RNA template allowing deletion of IESs and linking of MDSs in correct order

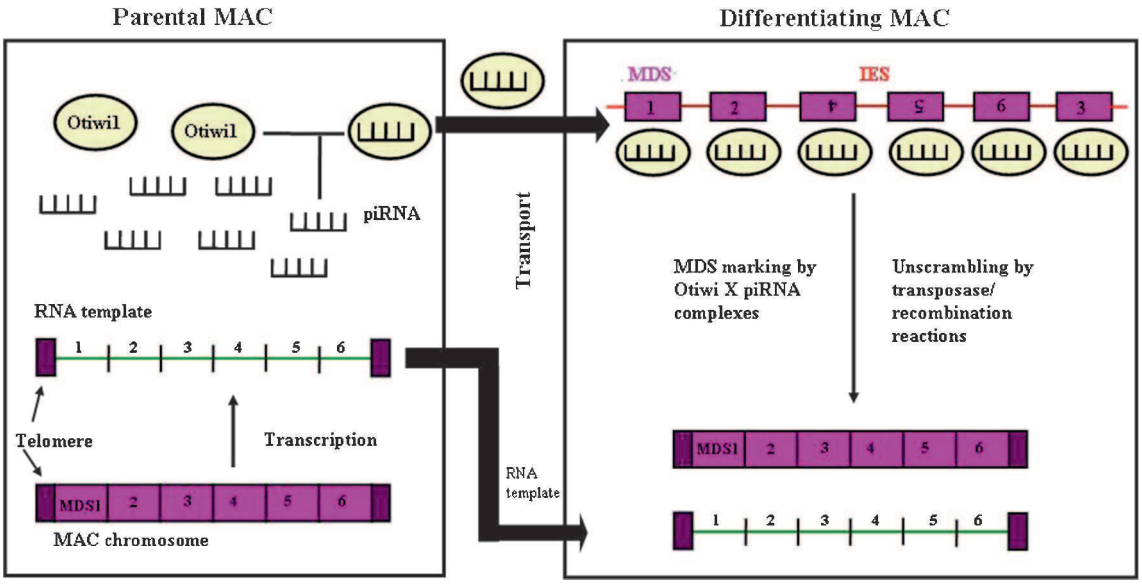

Figure 3. Hypothesized mechanism for generation of nanochromosomes during differentiation of micronucleus (MIC) into MAC soon after sexual reproduction. It is visualized that the locus comprised by MDS1-6 in the degrading parental MAC produces a telomere containing transcript, the template RNA. This RNA is processed into 27 nt piRNAs. The Piwi protein, Otiwi1 complexes with piRNAs. These Otiwi X piRNA complexes are transported into MIC undergoing development into MAC. Here, the piRNAs in RNA X protein complexes recognize the complementary MDS sequences and mark (shield) them for retention during the process of DNA rearrangements that follows. The old/parental MAC also transfers the template RNA to the developing MAC. The template RNA guides correct ordering of MDSs 1–6, using the enzymatic system introduced in figure 2. The homology of a specific whole gene in the template RNA with the MDSs of the corresponding genes in DNA from MIC allows correct ordering of MDSs for their recombinational stitching. Any remaining pointer sequence at MDS junctions or errors are repaired by the DNA repair mechanism guided by the homology of the gene reconstructed from DMSs with template RNA. The gene product from DMSs free of any error is capped with telomeres to produce a nanochromosome competent to specify a gene product during asexual growth and development. The nanochromosome undergoes 1000 or more folds amplification as shown in figure 2.

comes from two transgenosis experiments. Transgenic introduction of a synthetic long RNA carrying altered sequence of MDSs resulted in production of correspondingly aberrant form of nanochromosome which gets transgenerationally propagated. Transgenic introduction of synthetic piRNAs homologous to regions that are normally deleted led to their retention and the aberrantly retained sequences continued to be present over several generations (Nowacki *et al*. 2008, 2009; Nowacki and Landweber 2009; Nowacki *et al*. 2011; Fang *et al.* 2012; Bracht *et al.* 2013).

Involvement of RNA copies of nanochromosomes of the parental MAC from previous sexual generation provides a mechanism for transgenerational perpetuation of the population spectrum of nanochromosomes of all the vegetative generations proceeding to the last sexual reproduction.

Functional advantages rendered by the programmed genome rearrangement

O. trifallax is commonly found in ponds across continents. Its success may be related to several advantages accruing to its populations from programmed genome rearrangement of its somatic nucleus which are as follows: (i) removal of all extra DNA allows developmental stage-specific gene expression. (ii) Gene/nanochromosome amplification permits high gene expression levels. (iii) High levels of ploidy serve as buffer for the loss of chromosomes during divisions on account of absence of centromeres from chromosomes. (iv) Differential amplification of nanochromosomes and their alternate forms increase variability in population. (v) Presence of ploidy increase chances of new alleles arising from errors in DNA replication. (vi) Recombination between distinct nanochromosomes bearing regions of homology gives rise to new nanochromosomes, thereby increasing genetic variability. The analysis of single-nucleotide polymorphism (SNP) in its isolates has revealed that *O. trifallax* may be one of the most populous living organisms (Finlay *et al.* 1996; Zoller *et al.* 2012; Chen *et al.* 2014).

Evolutionary significance of programmed genome editing

Higher eukaryotic plant and animal genomes are often laden with transposons; about 50% *O. trifallax* MIC genome comprises transposons and other such sequences. *O. trifallax* uses transposons to provide transposase activity in rearranging its somatic nucleus, in MACs undergoing differentiation; however, these get deleted subsequently such that MAC becomes free of transposon presence. In higher plants and animals, the transposons are kept silenced by heavy cytosine methylation

and/or chromatin heterochromatization. Apparently, the two pathways have evolved to control the coevolving transposons in genome expansions. The pathway of *O. trifallax* is inapplicable in higher plants and animals as their cells have only one nucleus that participates both in sexual reproduction and asexual cell divisions during vegetative/somatic growth of organs. In higher plants and animals, programmed rearrangement of gene segments, such as in alternate forms of nanochromosomes arising from faulty fusion of MDSs in *O. trifallax*, may allow construction of gene variants to perform similar type of multiple functions as exemplified by antibody variability arising from combinations of V, D and J genes in humans (Market and Papavasiliou 2003). This phenomenon awaits discovery in higher plants. Thus, the mechanism of programmed gene rearrangement in higher eukaryotes has been retained with respect to certain specific cluster of genes during their evolution.

Some interesting questions posed by the ciliate nuclear dynamics

Table 1 compares some of the genomic properties of the four species in the protist clade alveolata: the ciliates, *O. trifallax*, *Stylonychia lemnae* and *Paramecium tetraurelia*, and the apicomplexan malarial parasite, *Plasmodium falciparum*. Adenine–thymine richness of the genome and use of next-generation sequencing (NSG) techniques complemented by single-molecule real time sequencing (SMRT) technique (Roberts *et al*. 2013; Swart *et al*. 2013; Aeschlimann *et al*. 2014) have enabled comprehensive genomic analyses in these organisms. It is noted that the genomes of *Oxytricha*, *Stylonychia* and *Paramecium* are more complex than the genome of *Plasmodium*. *Stylonychia* and *Paramecium* derive their MAC much like *Oxytricha*. MAC genome organization of *Oxytricha* and *Stylonychia* is similar and there is high degree of synteny between the nanochromosomes of these two ciliates. However, in *Paramecium*, only about 10–20% of its MIC genome undergoes programmed rearrangements, unlike *Oxytricha* and *Stylonychia*, in which 90–95% and 90% of MIC genome undergoes programmed rearrangements, respectively. The MIC and MAC of *Paramecium* contain about 50 and 200–350 chromosomes, respectively. Increase in chromosome copy number (ploidy) is also lower in *Paramecium*, about 800–1000 fold in *Paramecium* as compared with *Oxytricha*, which is about 2000 fold and *Stylonychia* about 1500 fold. The micronucleus/macronucleus of their sister alveolate *Plasmodium* contains only two copies of 14 chromosomes. The relative fragmentation of the MAC genome (or chromosome complement size) in four species is estimated as, *Plasmodium* : *Paramecium* : *Stylonychia* : *Oxytricha* :: $1: 4 \times 10^3 : 4 \times 10^5 : 5 \times 10^5$.

The occurrence of genetic variability in the designs of MAC/MIC among alveolates offer opportunity to discover alleles in *O. trifallax* and *P. tetraurelia* that are involved in the various steps of MAC differentiation from MIC mother and those that determine genome fragmentation and polyploidization of the nanochromosomes in the MAC,

Table 1. Some properties of the MIC and MAC genomes of *Plasmodium falciparum*, *Paramecium tetraurelia*, *Stylonychia lemnae* and *Oxytricha trifallax* alveolate.

Designation/ serial number	Nuclear property	Ciliate			Apicomplexan
		Oxytricha trifallax	*Stylonychia lemnae*	*Paramecium tetraurelia*	*Plasmodium falciparum*
A	MIC				
1	Genome size (Mb)	490–500	500	100–120	23
2	Chromosome number (*n*)	100	100	50	14
3	Ploidy (= *xn*)	2	2	2	2
B	MAC				
4	Genome size (Mb)	50	49	72	23
5	Chromosome number (*n*)	15600	16000	200	14
6	Ploidy (= *xn*)	1900	1500	800	2
7	Alternate chromosomal fragmentation	Yes	Yes	Limited	No
8	Degree of synteny of nanochromosomes with those of *O. trifallax*	NA	High	NK	None
9	Base composition (A + T per cent)	66	68	72	82
References		Lauth *et al.* (1976); Raikov (1982); Swart *et al.* (2013); Chen *et al.* (2014)	Steinbruck (1983); Xu *et al.* (2012); Swart *et al.* (2013); Aeschlimann *et al.* (2014)	Zagulski *et al.* (2004); Aury *et al.* (2006); Arnaiz and Sperling (2011); Swart *et al.* (2013)	Weber (1987); Gardner *et al.* (2002); Swart *et al.* (2013)

NA, not applicable; NK, not known.

which are either absent or not functionally conducive in *Plasmodium falciparum*. In the first instance, there is need to identify among the genes that are exclusively expressed in MIC and those that regulate MIC to MAC transition (or MAC differentiation). There is also scope to find in the ciliate model systems how the following phenomena occur: (i) the MAC and MIC vegetative divisions are synchronized and the integrities of the two kinds of nuclei are maintained; (ii) the internuclear transfer of RNA is promoted and interfered; (iii) the functional completeness is achieved in genes, postrecombinational stitching at pointers; (iv) the copy number of different nanochromosomes/chromosomes is controlled; (v) the gene expression from many copies of a gene lying on independent chromosomes gets synchronized or controlled; (vi) the homologous chromosomes get arranged with respect to each other and other sets of chromosomes in the interphase of cell division cycle; (vii) the chromatids get separated in the amitotic divisions and (viii) the role of chromosome number in the determination of nuclear size and in turn the size of cytoplasm and number and sizes of nonnuclear organelles. The *O. trifallax* and *P. tetraurelia* on one hand and the *Plasmodium falciparum* system on the other hand are suitable to seek answers to a plethora of questions about the known biological differences between them. In this regard, forward and reverse genetics using random and directed mutagenesis, RNAi technique and transgenosis studies will prove highly fruitful. Several important new biological principles are likely to emerge from the comparative studies of *Oxytricha*, *Paramecium* and *Plasmodium* alveolates. It is surmised that such analyses will advance our understanding of eukaryotic biology in general.

Acknowledgement

SK expresses grateful thanks towards Indian National Science Academy for the award of an honorary scientistship.

References

Aeschlimann S. H., Jönsson F., Postberg J., Stover N. A., Petera R. L., Lipps H. J. *et al.* 2014 The draft assembly of the radically organized *Stylonychia lemnae* macronuclear genome. *Genome Biol. Evol.* **6**, 1707–1723.

Arnaiz O. and Sperling L. 2011 Paramecium D. B. in 2011: new tools and new data for functional and comparative genomics of the model ciliate *Paramecium tetraurelia*. *Nucleic Acids Res.* **39**, D632–636.

Aury J. M., Jaillon O., Duret L., Noel B., Jubin C., Porcel B. M. *et al.* 2006 Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444**, 171–178.

Bracht J. R., Fang W., Goldman A. D., Dolzhenko E., Stein E. M. and Landweber L. F. 2013 Genomes on the edge: programmed genome instability in ciliates. *Cell* **152**, 406–416.

Chen X., Bracht J. R., Goldman A. D., Dolzhenko E., Clay D. M., Swart E. C. *et al.* 2014 The architecture of a scrambled genome reveals massive levels of genomic rearrangement during development. *Cell* **158**, 1187–1198.

Fang W. and Landweber L. F. 2013 RNA-mediated genome rearrangement: hypotheses and evidence. *Bioessays* **35**, 84–87.

Fang W., Wang X., Bracht J. R., Nowacki M. and Landweber L. F. 2012 Piwi-interacting RNAs protect DNA against loss during *Oxytricha* genome rearrangement. *Cell* **151**, 1243–1255.

Finlay B. J., Maberly S. C. and Esteban G. F. 1996 Spectacular abundance of ciliates in anoxic pond water: contribution of symbiont photosynthesis to host respiratory oxygen requirements. *FEMS Microbiol. Ecol.* **20**, 229–235.

Gardner M. J., Hall N., Fung E., White O., Berriman M., Hyman R. W. *et al.* 2002 Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511.

Lauth M. R., Spear B. B., Heumann J. and Prescott D. M. 1976 DNA of ciliated protozoa: DNA sequence diminution during macronuclear development of Oxytricha. *Cell* **7**, 67–74.

Market E. and Papavasiliou F. N. 2003 V(D)J recombination and the evolution of the adaptive immune system. *PLoS Biol.* **1**, e16.

Nowacki M. and Landweber L. F. 2009 Epigenetic inheritance in ciliates. *Curr. Opin. Microbiol.* **12**, 638–643.

Nowacki M., Vijayan V., Zhou Y., Schotanus K., Doak T. G. and Landweber L. F. 2008 RNA-mediated epigenetic programming of a genome-rearrangement pathway. *Nature* **451**, 153–158.

Nowacki M., Higgins B. P., Maquilan G. M., Swart E. C., Doak T. G. and Landweber L. F. 2009 A functional role for transposases in a large eukaryotic genome. *Science* **324**, 935–938.

Nowacki M., Haye J. E., Fang W., Vijayan V. and Landweber L. F. 2010 RNA-mediated epigenetic regulation of DNA copy number. *Proc. Natl. Acad. Sci. USA* **107**, 22,140–22,144.

Nowacki M., Shetty K. and Landweber L. F. 2011 RNA-mediated epigenetic programming of genome rearrangements. *Annu. Rev. Genomics Hum. Genet.* **12**, 367–389.

Raikov I. B. 1982 *The protozoan nucleus*. Springer-Verlag, Vienna, Austria.

Roberts R. J., Carneiro M. O. and Schatz M. C. 2013 The advantages of SMRT sequencing. *Genome Biol.* **14**, 405.

Steinbruck G. 1983 Overamplification of genes in macronuclei of hypotrichous ciliates. *Chromosoma* **88**, 156–163.

Swart E. C., Bracht J. R., Magrini V., Minx P., Chen X., Zhou Y. *et al.* 2013 The *Oxytricha trifallax* macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. *PLoS Biol.* **11**, e1001473.

Weber J. L. 1987 Analysis of sequences from the extremely A +T-rich genome of *Plasmodium falciparum*. *Gene* **52**, 103–109.

Xu K., Doak T. G., Lipps H. J., Wang J., Swart E. C. and Chang W. J. 2012 Copy number variations of 11 macronuclear chromosomes and their gene expression in *Oxytricha trifallax*. *Gene* **505**, 75–80.

Zagulski M., Nowak J. K., Le Mouel A., Nowacki M., Migdalski A., Gromadka R. *et al.* 2004 High coding density on the largest *Paramecium tetraurelia* somatic chromosome. *Curr. Biol.* **14**, 1397–1404.

Zoller S. D., Hammersmith R. L., Swart E. C., Higgins B. P., Doak T. G., Herrick G. and Landweber L. F. 2012 Characterization and taxonomic validity of the ciliate *Oxytricha trifallax* (Class Spirotrichea) based on multiple gene sequences: limitations in identifying genera solely by morphology. *Protist* **163**, 643–657.

Received 4 October 2014, in revised form 22 November 2014; accepted 1 December 2014
Unedited version published online: 18 December 2014
Final version published online: 4 June 2015