RESEARCH ARTICLE

Genomic Characterization of Variable Surface Antigens Reveals a Telomere Position Effect as a Prerequisite for RNA Interference-Mediated Silencing in *Paramecium tetraurelia*

Damir Baranasic,[c,d] Timo Oppermann,[b] Miriam Cheaib,[a,b] John Cullum,[c] Helmut Schmidt,[b] Martin Simon[a]

Saarland University, Centre for Human and Molecular Biology, Molecular Cellular Dynamics, Saarbrücken, Germany[a]; Department of Biology, University of Kaiserslautern, Kaiserslautern, Germany[b]; Department for Genetics, Faculty of Biology, University of Kaiserslautern, Kaiserslautern, Germany[c]; Faculty of Food Technology and Biotechnology, University of Zagreb, Zagreb, Croatia[d]

ABSTRACT **Antigenic or phenotypic variation is a widespread phenomenon of expression of variable surface protein coats on eukaryotic microbes. To clarify the mechanism behind mutually exclusive gene expression, we characterized the genetic properties of the surface antigen multigene family in the ciliate *Paramecium tetraurelia* and the epigenetic factors controlling expression and silencing. Genome analysis indicated that the multigene family consists of intrachromosomal and subtelomeric genes; both classes apparently derive from different gene duplication events: whole-genome and intrachromosomal duplication. Expression analysis provides evidence for telomere position effects, because only subtelomeric genes follow mutually exclusive transcription. Microarray analysis of cultures deficient in Rdr3, an RNA-dependent RNA polymerase, in comparison to serotype-pure wild-type cultures, shows cotranscription of a subset of subtelomeric genes, indicating that the telomere position effect is due to a selective occurrence of Rdr3-mediated silencing in subtelomeric regions. We present a model of surface antigen evolution by intrachromosomal gene duplication involving the maintenance of positive selection of structurally relevant regions. Further analysis of chromosome heterogeneity shows that alternative telomere addition regions clearly affect transcription of closely related genes. Consequently, chromosome fragmentation appears to be of crucial importance for surface antigen expression and evolution. Our data suggest that RNAi-mediated control of this genetic network by *trans*-acting RNAs allows rapid epigenetic adaptation by phenotypic variation in combination with long-term genetic adaptation by Darwinian evolution of antigen genes.**

IMPORTANCE **Alternating surface protein structures have been described for almost all eukaryotic microbes, and a broad variety of functions have been described, such as virulence factors, adhesion molecules, and molecular camouflage. Mechanisms controlling gene expression of variable surface proteins therefore represent a powerful tool for rapid phenotypic variation across kingdoms in pathogenic as well as free-living eukaryotic microbes. However, the epigenetic mechanisms controlling synchronous expression and silencing of individual genes are hardly understood. Using the ciliate *Paramecium tetraurelia* as a (epi)genetic model, we showed that a subtelomeric gene position effect is associated with the selective occurrence of RNAi-mediated silencing of silent surface protein genes, suggesting small interfering RNA (siRNA)-mediated epigenetic cross talks between silent and active surface antigen genes. Our integrated genomic and molecular approach discloses the correlation between gene position effects and siRNA-mediated *trans*-silencing, thus providing two new parameters for regulation of mutually exclusive gene expression and the genomic organization of variant gene families.**

Received 1 June 2014 **Accepted** 24 June 2014 **Published** 11 November 2014

Citation Baranasic D, Oppermann T, Cheaib M, Cullum J, Schmidt H, Simon M. 2014. Genomic characterization of variable surface antigens reveals a telomere position effect as a prerequisite for RNA interference-mediated silencing in *Paramecium tetraurelia*. mBio 5(6):e01328-14. doi:10.1128/mBio.01328-14.

Editor Keith Gull, University of Oxford

Address correspondence to Martin Simon, martin.simon@uni-saarland.de

Programmed antigenic variation is the ability of eukaryotic microbes to alter the antigenic molecules exposed on their surfaces by mutually exclusive expression of a multigene family of variant surface antigens (1). This phenomenon has been observed in a wide range of pathogenic microorganisms, where it enables them to escape attack by the hosts' immune system. The free-living microbe *Paramecium tetraurelia* can also switch surface proteins. Early studies revealed the presence of different serotypes using antibodies raised by injection of paramecia into rabbits (2). The literature usually refers to the responsible proteins as surface antigens (SAgs), by analogy to the systems in related parasitic ciliates (3, 4), although a more precise term would be phenotypic variation. Several studies indicate that the phenotypic variation of SAgs may protect paramecia from predators, either by masking mechanisms or by jettisoning antigens during attacks (5, 6). It is not clear whether variation has other roles, as it can also be triggered by changing environmental factors, such as temperature or salinity, or changing biotic factors, such as availability of food or addition of homologous antibodies (7). Such phenotypic variation has been reported for many free-living fungi and protists as well as most pathogens, and common mechanisms seem to be involved for distantly related species (8). This makes it attractive

to use nonpathogenic model organisms such as *Saccharomyces cerevisiae* and *Paramecium tetraurelia* to study the genome organization and epigenetic mechanisms controlling antigenic variation (9, 10).

A variety of mechanisms have been implicated in antigenic variation in different species. A well-studied example involving DNA rearrangements is *Trypososnosma brucei*, where gene conversion events copy entire genes or fragments of genes into active expression sites; segmental gene conversions also create new chimeric genes, thus increasing antigenic variability (11). However, even in *T. brucei*, *in situ* switches can occur with epigenetic chromatin modifications instead of DNA sequence alterations (11, 12). In many organisms, *in situ* switches seem to be the rule, e.g., the highly variable erythrocyte membrane protein 1 of *Plasmodium falciparum* (PfEMP1) and the variant surface proteins (VSPs) of *Giardia lamblia*. Interestingly, the silencing of nonexpressed surface antigens occurs at different levels, as *P. falciparum* regulates PfEMP1 at the level of transcription, whereas *G. lamblia* transcribes all VSP genes and then posttranscriptionally cleaves mRNA of the nonexpressed genes using the RNA interference (RNAi) pathway (13, 14).

In many species (e.g., free-living and parasitic fungi as well as *P. falciparum* and *T. brucei*), surface antigen genes have a subtelomeric location. It has been suggested that spreading of telomeric heterochromatin into subtelomeric regions might result in the transcriptional silencing of surface antigen genes (15). In *P. falciparum*, chromatin remodeling is involved in antigenic variation of the *var* gene family, as silent loci show high levels of H3K9 trimethylation, whereas active loci are characterized by H3 acetylation (reviewed in reference 16). In some organisms, such as *T. brucei* and *P. falciparum*, activation and silencing of the genes in association with epigenetic chromatin modifications were shown to involve transport of the genes into distinct subnuclear compartments, which allow active transcription (12, 17). In contrast, the posttranscriptionally controlled *Giardia* antigen genes do not show preferred subtelomeric localization. An involvement of transcription-regulating small interfering RNAs (siRNAs) comparable to the posttranscriptional *Giardia* system has not been shown for any organism exhibiting a telomere position effect of its antigen genes.

The serotype of *Paramecium tetraurelia* is due to the expression of a single surface antigen gene (*SAg*), and serotypes are inherited in a non-Mendelian manner, meaning that the gene expression pattern becomes transmitted to progeny by unknown epigenetic mechanisms (18). Serotype switches can occur at any time without the need for sexual recombination. Individual serotype proteins show highly variable sequence areas in their central regions, indicating that diversity plays a major role in protein function. Comparison of the genes indicated that all these surface antigen proteins have an extraordinarily high percentage of cysteine which is distributed throughout the polypeptide with a highly specific periodicity (19, 20) which is similar to the VSG in *T. brucei* (21). Both organisms were also reported to use GPI (glycosylphosphatidylinositol) anchors to tether the surface antigen to the outer leaflet of the external membrane (22, 23). *Paramecium* surface antigens share some similarity in their N- and C-terminal regions but have highly variable regions in the center (4, 19). Analysis of the ratio of nonsynonymous to synonymous substitutions (dN/dS ratio) showed a low proportion of nonsynonymous base changes in the terminal regions of the genes, indicating the action of purifying selection to conserve the amino acid sequence, whereas the central areas show high ratios of nonsynonymous substitutions (24). Interestingly, these central regions often consist of internal tandem repeats. This suggested that the central region was responsible for the diversity of the surface proteins, which would be selected for during evolution, whereas the conserved terminal regions might have roles in maintaining tertiary structure and membrane anchoring. Immunological analyses support this hypothesis, indicating that the variable central parts are exposed to the medium, whereas the N- and C-terminal areas are immunologically hidden close to the membrane (25).

We demonstrated that *Paramecium* serotypes are controlled by RNA interference (26). Knockdown of an RNA-dependent RNA polymerase (Rdr3) resulted in a stable coexpression of all tested surface antigens rather than the exclusive expression of one surface antigen, as usually occurs during antigenic variation. This was accompanied by the disappearance of endogenous siRNAs. In contrast to the report of RNAi-controlled VSP regulation in *Giardia*, small RNAs seem to regulate *Paramecium* serotypes at a transcriptional level rather than through mRNA stability.

In this study, we used bioinformatics methods to characterize the genetic properties of the *SAg* multigene family. Surprisingly, most of the genes in this family are not ohnologous genes, and moreover, several of them show gene duplicates located on the same macronuclear chromosome. Because of three successive whole-genome duplications, *Paramecium tetraurelia* became a model to study the parameters forcing gene duplicate retention and loss (27), allowing us to interpret our data on *SAg* ohnologs and other gene duplicates in context with the requirements for subtelomeric localization to build a model for serotype evolution, which includes the crucial parameters for RNAi-mediated, mutually exclusive expression of *SAg* genes described here.

RESULTS

The *SAg* gene family encodes large surface membrane-located proteins. Eleven serotypes have been described for *P. tetraurelia* using immunological methods, but until recently, sequence data were available only for genes corresponding to 6 serotypes: *51A* (28), *51B* (29), *51C* (20), *51D* (30), *51G* (31), and fragments of *51H* (32). Although the limited available sequence data for *Paramecium* serotypes indicate some conserved sequences in the 5′ and 3′ coding regions of genes (20), this was not sufficient to allow a similarity-based database mining approach. Therefore, we used another exclusive feature of these proteins to identify further *SAg* genes in the genome: all serotypes sequenced so far showed a highly conserved cysteine periodicity throughout the protein, and we used a Pfam domain that was built accordingly for a proteome-wide *in silico* search (*Paramecium* peptides at *Paramecium*DB [33]). We extracted 96 candidate serotypes from the proteome with intact cysteine periodicity (Pfam *E* value, $\leq 10^{-9}$); Fig. S1A in the supplemental material shows the cysteine periodicity of 51Hβ and SAg 94, representative of all other proteins.

To decide whether genes have intact open reading frames, we manually reassembled the sequencing reads of the genome project. This was done to correct errors of the automated assembly and to verify start and stop codons by successive translation verifying N-terminal endoplasmic reticulum (ER) translocation signals and C-terminal GPI-anchoring peptides or transmembrane domains (see below). This was possible for 65 genes. Because of the high similarity of the 5′ and 3′ coding regions and because of a low

coverage of loci on very tiny macronuclear chromosomes, a complete reassembly of 31 of the genes failed. It is thus not clear whether these are intact (protein coding) genes. The subsequent analyses were confined to the 65 correctly reassembled genes; see Table S1 in the supplemental material for details.

The 65 verified sequences were analyzed in more detail. Sixty of the proteins had N-terminal ER translocation signals (ER-TLS), suggesting that the proteins entered the secretory pathway (see Fig. S2 in the supplemental material). Of these, 55 SAgs contained C-terminal signal peptides predicting GPI anchors (see Fig. S3 in the supplemental material). One of these genes (*SAg 83*) was probably a pseudogene, because of the presence of a stop codon in the predicted reading frame. Thus, there were 54 genes with the 8-amino-acid cysteine periodicity, which could encode surface proteins with a GPI anchor. These included the four previously described sequences (*51B*, *51C*, *51D*, and *51G*) as well as the complete sequence of the *51H* gene (only segments were published earlier). In addition, the serotype genes *51I* and *51J* were identified using the unpublished sequences of cDNA fragments (A. Valatka and H. Schmidt, personal communication). The genome sequence used strain *P. tetraurelia* d4-2, which has the genetic background of wild-type stock 51 but has the *51A* serotype allele replaced by the *29A* allele from stock 29. Thus, genes corresponding to 8 of the 11 known serotypes could be identified.

The gene sequences were also analyzed for the presence of potential transmembrane domains. In eight cases, C-terminal transmembrane domains were predicted. These proteins did not contain signal peptides for GPI anchors, suggesting that there might be an alternative mechanism for anchoring SAgs. Three of the genes contained ER-TLS, suggesting that the proteins would enter the secretory pathway, but the other five genes did not contain a detectable ER-TLS.

Putative *SAg* genes are large (the average length of the coding region is 6.6 kb, compared to an average of 1.4 kb for all protein-coding genes of *P. tetraurelia*), which is in agreement with the sizes of previously sequenced *SAg* genes (see Fig. S4A in the supplemental material). However, most of the *SAg* genes do not contain introns (four genes each contain a single intron), whereas *P. tetraurelia* protein-coding genes contain an average of 2.9 introns. The G+C content (34.4%) is also higher than that of the average gene (30.4%) (see Fig. S4B in the supplemental material). There are four large superantigen genes with an average coding region of about 12 kb.

Serotype isoforms result from intrachromosomal gene duplications. Figure 1A shows a neighbor joining tree of proteins indicating three separated phylogenetic clusters, as indicated by the background shading: all proteins with transmembrane domains, a cluster containing the 51D and 51H serotypes with their individual isoforms (see below), and the largest cluster, which includes the 29A, 51B (and isoforms), 51G, 51C, and 51I serotype proteins as well as the superantigens. Only SAg 3 cannot be related to one of the clusters. This clustering is consistent with previous immunological studies describing serotype proteins 51A, 51B, 51G, and 51Q as well as a group immunologically related to serotypes 51D and 51J (34), although the gene for serotype 51Q remains to be annotated.

Next to a close relationship between individual serotype proteins, e.g., proteins 51D and 51J or proteins 51A, 51B, and 51G, the tree shows isoforms for the genes encoding 51B, 51D, and 51H. These show a very high degree of identity, including 91% identity between 51Hα and 51Hβ, an average identity of 79% for the four 51D isoforms, and an average identity of 82% for the four 51B isoforms, so their identity is drastically higher than the average for all SAgs (~27% identity). Although isogenes of the *51D* gene were reported earlier (35) this was not reported for any other serotype gene in *P. tetraurelia*. Because of their high similarity, these isogenes appear to result from very recent gene duplications.

Aury et al. described the sequencing of the macronuclear genome and its evolution by three successive whole-genome duplications (WGD); consequently, the genome consists of a high number of ohnologous gene duplicates (36). As mentioned above, serotype genes share an individual degree of homology, and therefore, it seemed tempting to speculate that gene duplication in the context of whole-genome duplication also contributed to serotype diversity. However, this is not the case, because our data indicate that only 25 ohnologs of the ancient and intermediary WGD exist, but none of the classically described surface antigen genes have ohnologs (Fig. 1A). The neighbor joining tree indicates that the transmembrane antigens (TMA) are separated, whereas SAgs showing ohnologous duplicates build a cluster in the center of the tree which is surrounded by small clusters of highly similar gene copies of the classical surface antigens. To analyze the chromosomal distribution of these genes, we mapped them to the genome assembly; most of the assembled scaffolds do indeed represent macronuclear chromosomes, as telomere reads can be found at both scaffold ends (36). Surprisingly, mapping of the antigen genes to the genome annotation revealed that almost all of the individual isogenes are located on the same macronuclear chromosome (Fig. 2). This is true for the four *51B* isogenes on scaffold 143 and two *51H* isogenes on scaffold 142. In the case of the *51D* isogenes, *51Dγ-1* and *51Dγ-2* are located on scaffold 159, and only *51J*, which was originally described for the *51Dε* isogene (30), is located on chromosome 128. Therefore, serotype isoforms appear to result from intrachromosomal duplication events. In contrast to these isogenes, chromosome 51 harbors four totally different *SAg* genes (*51C*, *SAg 3*, *SAg 42*, and *51G*) which do not show any phylogenetic relationship. Most of the *SAg* genes do not show intrachromosomal isoforms, e.g., *29A* on chromosome 106 (as well as all other genes not included in Fig. 2).

Genomic maps identify a subset of *SAg* genes close to telomere repeats. During the analysis of the gene maps, our attention was attracted to the gene position, as several genes appeared to be located close to the end of scaffolds, e.g., *29A* and *51J* (Fig. 2). To subsequently characterize chromosomal distribution, we had to consider the heterogeneity of *Paramecium* macronuclear chromosomes. In ciliates, there are two types of nuclei, because they show separation of germline and soma despite the fact that paramecia are unicellular organisms: the germline is represented by the micronuclei, which are responsible for sexual transmission, and the large macronuclei are responsible for somatic gene expression. During conjugation, macronuclei are generated from a zygotic nucleus by a complex process of DNA rearrangements and amplification to ~800n. During these processes, the large micronuclear chromosomes become shorter macronuclear chromosomes by two DNA elimination processes (reviewed in reference 37). Both precise elimination of IESs (internal elimination sequences) and the imprecise elimination of repeat sequences are responsible for heterogeneity of the macronuclear chromosomes (38). In other words, the macronuclear chromosomes exist in versions of different lengths. The published assembly ignores minority telomeres to

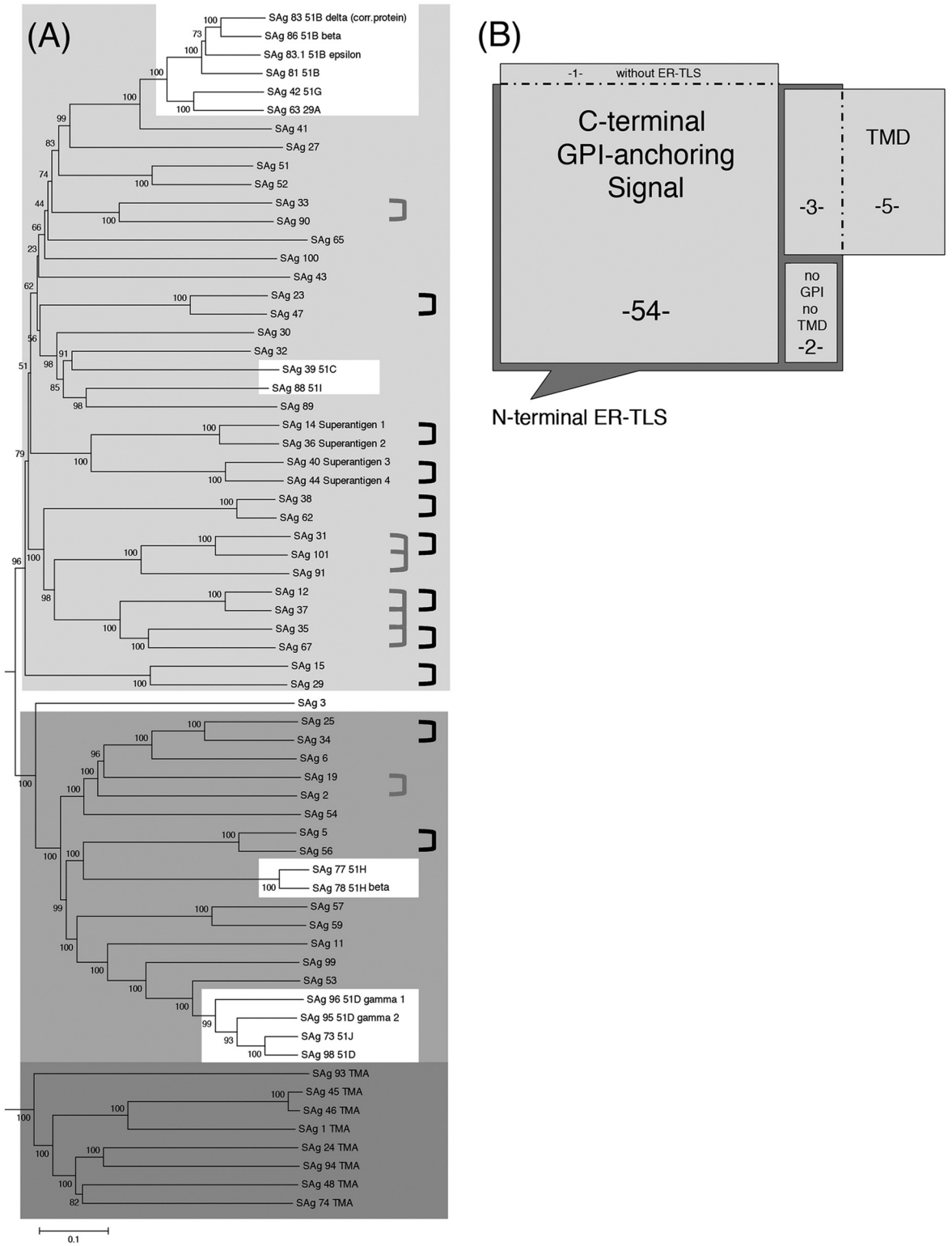

FIG 1 Evolutionary relationship between SAgs. (A) The neighbor joining tree (with 1,000 bootstrap replicates) is based on a multiple-sequence alignment of SAg proteins. The amino acid sequence of the defect SAg 83 (51Bδ) was corrected according to the 51Bα (SAg 81) sequence. Previously described SAgs and their newly identified isoforms are highlighted by the white background. Ohnologs of the whole genome duplications are indicated by black (most recent WGD) and gray (intermediate WGD) brackets. Bootstrap support values are given above the nodes; the scale bar indicates evolutionary distances. The three clusters of proteins discussed in the text are indicated by different background shading. (B) Schematic overview of the described proteins in terms of N-terminal endoplasmic reticulum translocation sequence (ER-TLS) and C-terminal anchoring prediction, GPI-anchoring, or transmembrane domains (TMD). Numbers represent proteins belonging to an individual group. The corrected sequence for SAg 83 was used.

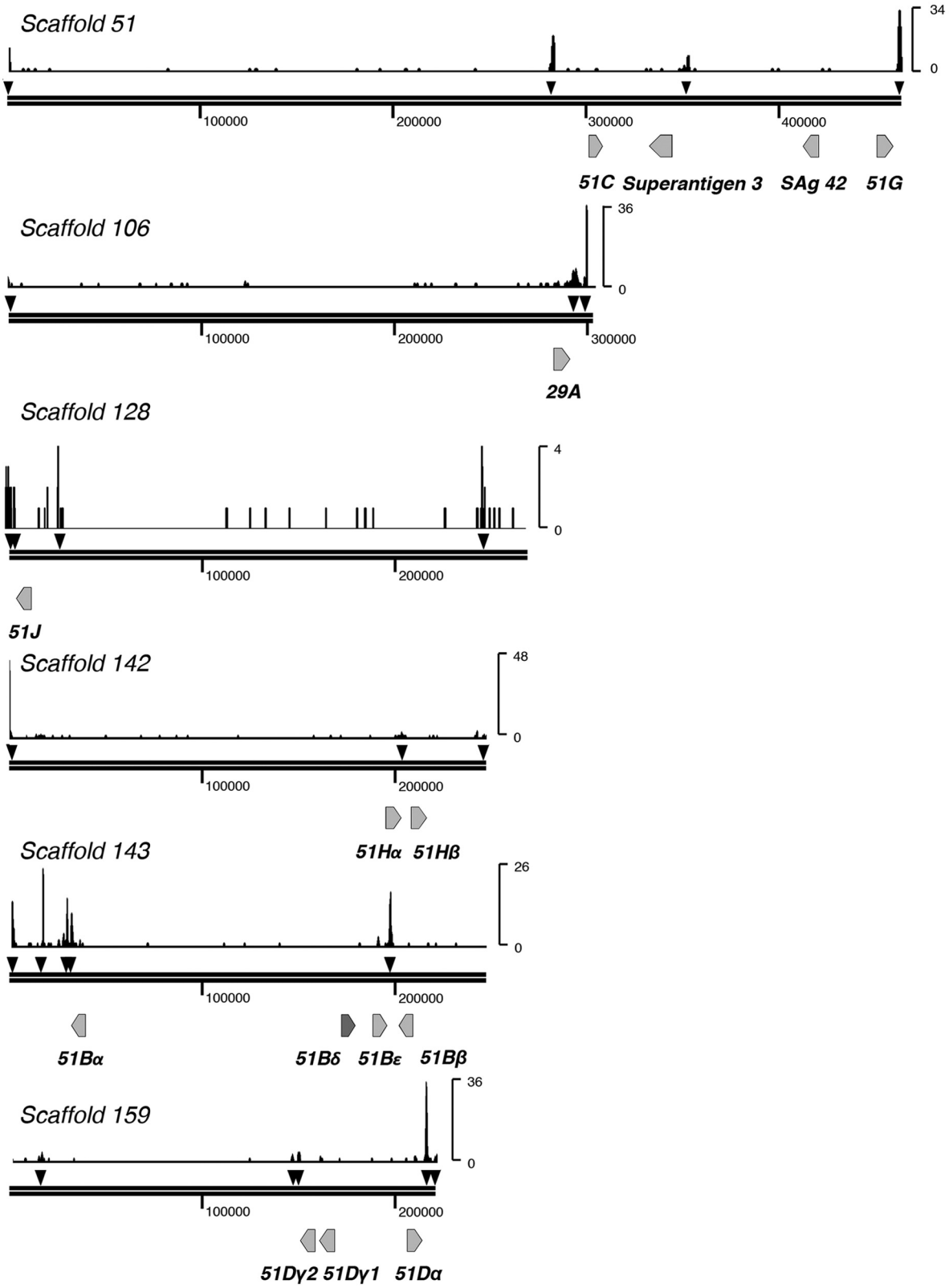

FIG 2 Chromosomal localization of individual surface antigen genes. Arrows indicate the positions and orientations of *SAg* genes on the individual macronuclear scaffold, represented by black double bar. Numbers below the bars are sizes, in bp. Intact genes are in gray, and the *51Dδ* pseudogene is in dark gray. To account for macronuclear heterogeneity, the genomic part of telomeric repeats containing reads was mapped to the genome. Arrowheads show telomeric sites, defined by a minimum of four overlapping reads. The maximum read count of telomere repeats containing reads is indicated on the right.

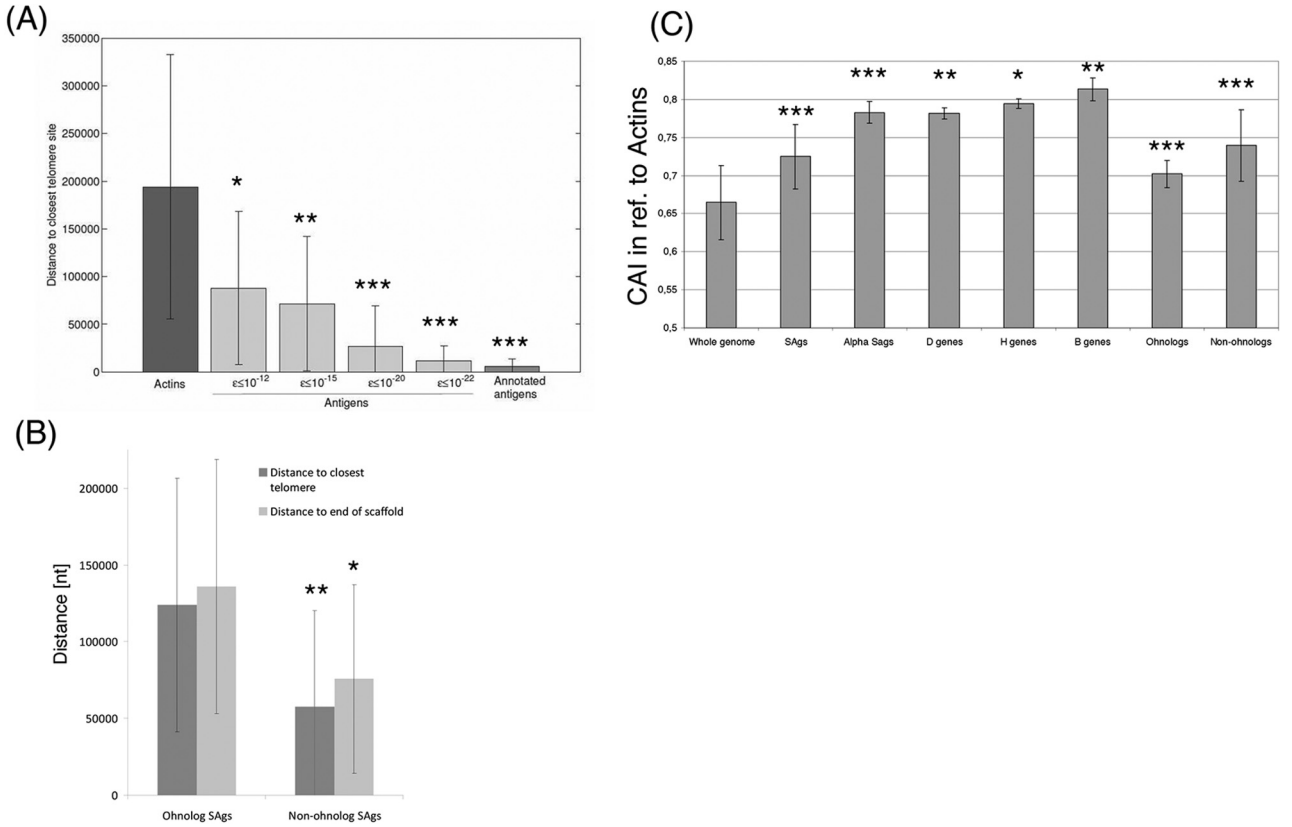

FIG 3 Analysis of chromosomal distribution of *SAg* genes in context of chromosome heterogeneity and whole-genome duplication. (A) Groups of SAgs defined by their *E* value to Pfam domain PF01508 indicating conformance with the eight-period cysteine periodicity of *Paramecium SAg* genes are set in relation to the mean distance of their genes to telomeric repeats, in bp. Lower *E* values indicate higher integrity of the cysteine periodicity. For the group of annotated antigens, the genes of *SAg 29A*, *51B*, *51C*, *51D*, *51H*, *51I*, and *51J* were selected. Error bars indicate standard deviations. Statistical analysis was done in reference to actins. The actin multigene family was included as a randomly distributed gene family. For distance calculation, also the macronuclear heterogeneity was considered, meaning that the distance to the closest mapped telomeric site was calculated. (B) Mean distance of ohnolog *SAg* genes and nonohnolog *SAg* genes to the closest telomeric site (dark gray) and to the ends of scaffolds (light gray). Error bars indicate standard deviations. (C) The codon adaptation index (CAI) was calculated for different gene sets in relation to *Paramecium* actin multigene family as a reference for highly expressed genes (whole genome = all predicted *Paramecium* CDS; *SAgs* = all 65 *SAg* genes; alpha *SAg* genes = *29A*, *51Bα*, *51C*, *51Dα*, *51G*, *51Hα*, *51I*, and *51J*; D genes = *51Dα*, *51Dγ-1*, and *51Dγ2*; H genes = *Hα* and *Hβ*; B isogenes = *Bα*, *Bβ*, *Bδ*, and *Bε*). All values were compared with the whole-genome CAI mean using a one-sample *t* test. If not stated otherwise, for the analyses testing the equality of the mean distances, the *t* test assuming unequal variances and unequal sample size was performed (*, $P \leq 0.05$; **, $P \leq 0.001$; ***, $P \leq 0{,}0001$); thus, we rejected the null hypothesis of equal means at a 5% error rate.

obtain long scaffolds, so that the question of whether a particular surface antigen gene has a subtelomeric location cannot always be answered from the assembled genome sequence. To characterize *SAg* genes in the genomic context, we took advantage of the whole-genome sequence (36) data and extracted all reads from the genome project containing at least three telomeric repeats (CCC[CA]AA) and allowing a single mismatch. After removal of these repeats, the genomic part of the reads was mapped to the genome assembly similar to the approach described in reference 38. To confirm the position of the genomic part of these reads, we examined the linking information of the paired-end sequencing by extraction of read mates and their subsequent mapping (data not shown). We defined a telomeric site as a minimum of four overlapping reads within a range of 2 kb, and these telomeric sites together with the *SAg* genes are shown in Fig. 2. For our prediction of macronuclear heterogeneity, the literature reports molecular evidence by restriction mapping or pulsed-field electrophoresis for telomeric sites on scaffolds 51 and 106 (39, 40), which supports our approach and the further analysis of genome-wide telomeric sites. Surprisingly, telomeric sites often appeared close to *SAg* genes, and this was especially the case for the annotated antigens. To quantify this effect, we calculated the distance of the individual genes to the closest telomeric site and set this in relation to the *E* value of the alignment with the domain profiles from Pfam, thus representing the integrity of the cysteine periodicity of the individual genes. Figure 3A shows that the mean distance of the gene to telomeres increases drastically for proteins showing a less conserved cysteine periodicity (indicated by increasing Pfam *E* values). In fact, these genes show a more or less uniform distribution in the genome, comparable to that of the actin multigene family, which was used as a control. *SAg* genes encoding proteins with intact cysteine periodicity, reflected by low *E* value, e.g., the annotated antigens (SAgs 29A, 51Bα, 51C, 51Dα, 51Hα, 51I, and 51J), show close proximity to telomeric sites. This would indicate that selection pressure occurs on cysteine periodicity in subtelomeric regions, which would require telomere-dependent expression. However, to distinguish whether a correlation between ohnologs and nonohnolog *SAg* genes can also be identified, Fig. 3B plots the

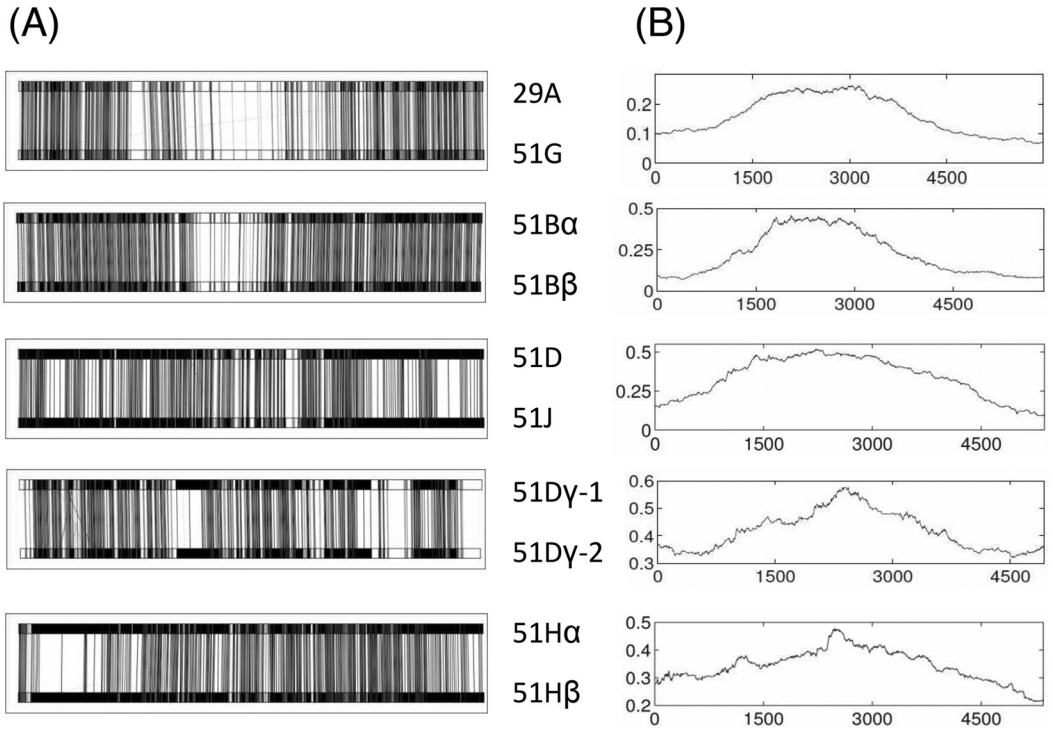

FIG 4 Comparative sequence analysis of SAg isoforms. (A) Pairwise alignments of the *SAg* genes indicated on the left. (B) dN/dS ratio calculated from the gene pairs shown in panel A using a sliding window of 450 to 1,000 bp and plotted against the position in the open reading frame.

mean distance of these two groups to either the closest telomeric site or the end of the scaffold, indicating a significantly closer proximity of the nonohnologous *SAg* genes to telomeres. It therefore seems tempting to speculate that gene duplication and gene position are important factors in evolution of distinct classes of *SAg* genes: subtelomeric and randomly distributed. Our data indicate that the two classes of *SAg* genes (subtelomeric and intrachromosomal) derive from different gene duplication mechanisms.

Positive selection pressure acts on N- and C-terminal protein areas. Any further interpretation of the two factors of gene duplication mechanism and subtelomeric localization in the context of serotype evolution remains speculative without gene expression data for isoforms and ohnologs, which would allow characterization of the telomere position effect on individual gene classes. To get a first indication of whether the individual isoforms are under selective pressure for codon optimization similar to that of highly expressed genes, we compared the codon adaptation index (CAI) using the highly expressed actin gene family to calculate the relative codon usage table. Figure 3C shows the CAI (calculated in reference to the highly expressed actin gene family) for the whole genome and the alpha *SAg* genes (encoding classical antigens) as well as the remaining antigens and isogene families. The significantly higher CAI of all *SAg* genes in comparison to the whole genome indicates their adaptation for high expression levels. Figure 3C also shows that the classical *SAg* genes (alpha genes and their isogenes) have a higher CAI than the entire *SAg* family. This would indeed indicate that the classical *SAg* genes and their isogenes differ from the others not only in subtelomeric localization but also in adaptation to high expression levels because of different expression behavior. In terms of the isogenes, it is not clear whether the elevated CAI is due to a short time distance to duplication or to a similar expression pattern to the alpha *SAg* genes.

Until now, the comparison of the available sequence data of different alleles of the same gene as well as alignments of different serotype proteins has indicated conservative N-terminal and C-terminal areas of the proteins, while variable areas can be identified in the central region (reviewed in reference 4). The variable central regions were moreover assumed to contain the immunological relevant information; thus, the distribution of synonymous mutations in N- and C-terminal areas and the accumulation of nonsynonymous mutations in the central areas are in agreement with a diversification of immunological information (24). This can also be seen in the comparison of the *29A* and *51G* genes, which shows variable areas in the center region (alignment in Fig. 4A) and relatively low nonsynonymous mutations in the marginal regions (Fig. 4B). In general, this distribution can be observed for all other comparisons of genes with other *SAg* genes or their individual isoforms. Of course, the degree of similarity and, as a prerequisite, the general level of substitutions vary, but a maximum dN/dS ratio exists in the central region for any analyzed gene pair (Fig. 4), suggesting that selection pressure acts on the marginal areas, which would subsequently indicate that isogenes are or were expressed to allow for selection pressure.

Intrachromosomal gene duplicates can be cotranscribed. Several independent studies reported the exclusive presence of a single *SAg* mRNA species in the *P. tetraurelia* surface antigen system (reviewed in reference 4). Classical Northern blot analysis using long double-stranded probes (Fig. 5A), however, cannot discriminate between different isoforms. We therefore analyzed

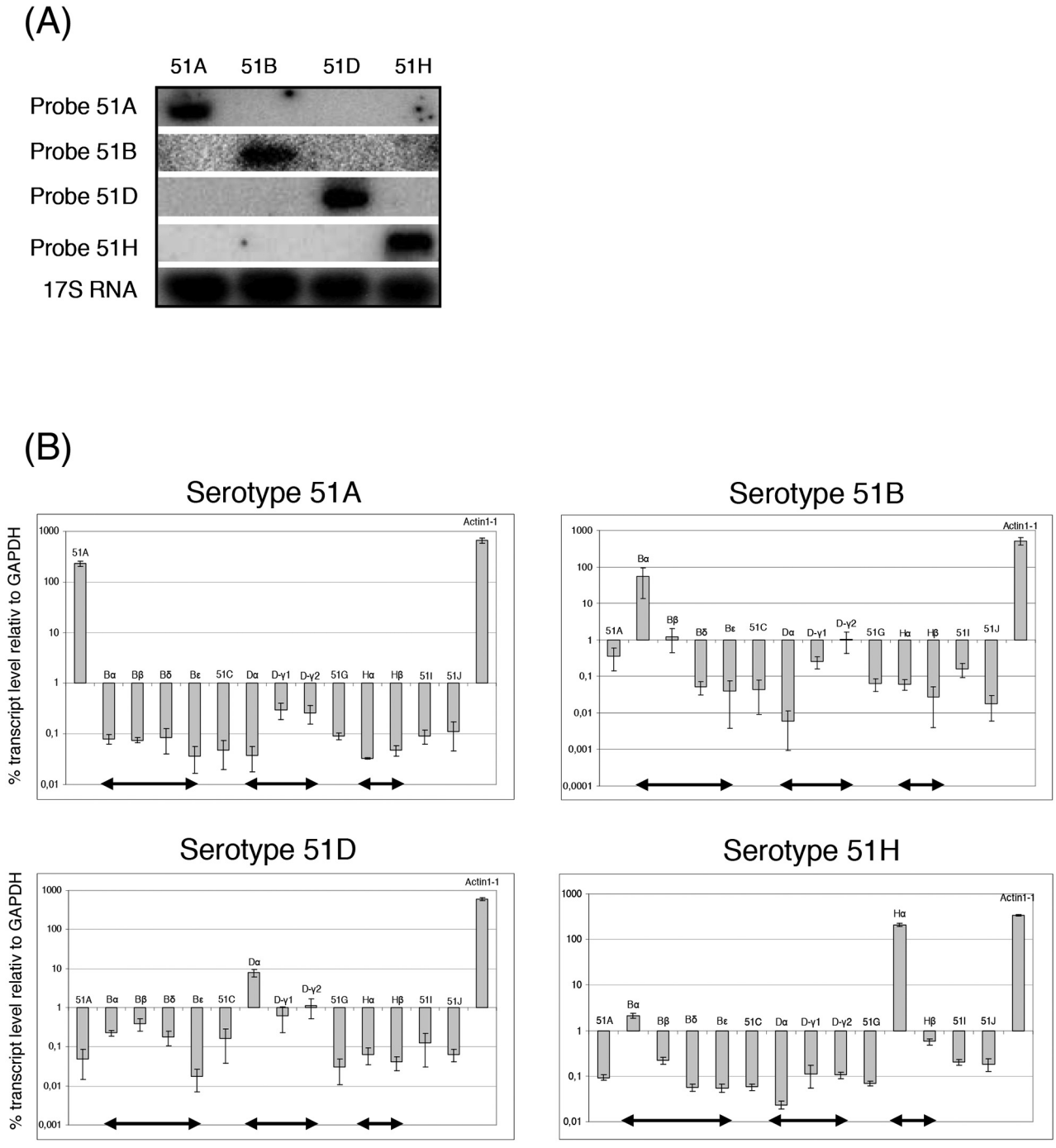

FIG 5 Analysis of *SAg* transcripts by Northern blotting and qPCR. (A) Northern blot of RNA isolated from serotypically pure 51A, 51B, 51D and 51H cultures hybridized with radiolabeled PCR product probes specific for the respective alpha-genes. 17S rRNA serves as a loading control (see Table S2 in the supplemental material for probe localization). (B) qRT-PCR screen of selected *SAg* genes and isogenes of the same RNA isolates used for Northern blots in >Fig. 4A (plus two additional biological replicates per serotype). The serotype of the culture in indicated on top of each graph. Data for individual genes were set in relation to transcript levels of the GAPDH gene by the ΔC_T method, and *actin1-1* served as a second housekeeping gene (see Table S1 in the supplemental material for individual qPCR product characteristics). Arrows indicate genes located on the same scaffold.

pure cultures of different serotypes for expression of isogenes using real-time PCR with primers designed to distinguish between isoforms. As shown in Fig. 5B, serotype 51A shows expression of a single antigen gene, *51A*. The pattern of serotype 51B is different: in addition to strong expression of the *51Bα* gene, the *51Bβ* gene shows transcription, but the isoforms *51Bδ* and *51Bε* do not; also, activation of the *51Dγ-2* gene can be observed (these cultures had a few contaminating 51A cells, explaining the slightly higher level of this gene). In serotype 51D, an upregulation of the *51Dα* gene as well as both gamma genes can be observed; however, the *51J* gene,

which shows a much higher similarity to *51D* than the gamma genes, does not show activation, indicating that cotranscription cannot be due solely to homology. Also, in serotype 51H, the beta isoform shows a slight upregulation together with strong activation of the *51Hα* gene. The 51H cultures were only 95% pure and had ~5% contaminating 51B cells; this can also be seen in the elevated values of the *51Bα* and the *51Bβ* genes.

These data indicate that isoforms of surface antigens can be cotranscribed with their respective alpha gene, although they never reach the same transcription level. This can be observed for serotypes 51B, 51D, and 51H, where the *51Bβ*, *51Dγ-1*, *51Dγ-2*, and *51Hβ* genes show also transcriptional activation. The fact that *51Dγ-1* as well as *51Bδ* and *51Dε* shows no or less transcriptional activation might indicate that this cotranscription occurs in a homology-dependent manner, as these genes are those with the lowest similarity to the alpha genes within the individual isogene families.

Considering a homology-dependent coactivation of isoforms, the comparison with other *SAg* genes shows that this homology-dependent activation and cotranscription occur only on isoforms located on the same chromosome (arrows in the graphs of Fig. 5B indicate genes located in the same macronuclear scaffold). This cotranscription does not occur if highly similar genes are located on different chromosomes: for instance, the *51J* gene, which shows the highest similarity to the *51Dα* gene, is not activated in serotype 51D cultures, and the *51G* gene does not show any cotranscription with the *51A* gene.

Interestingly, the phylogenetic relationship of the gene pairs 51A/51G and 51D/51J is closer than that of other isoforms: the difference that distinguishes them from other isoforms and integrates them in the mutually exclusive mechanism is apparently their location on a different chromosome.

Subtelomeric gene position is a prerequisite for RNAi-controlled mutually exclusive transcription. We previously reported that mutually exclusive expression of *SAg* genes in *Paramecium* is controlled by an RNAi pathway. A divergent RNA-dependent RNA polymerase (Rdr3) was shown to be involved in antigenic variation, because silencing of Rdr3 resulted in a phenotype of coexpression of different antigen genes. Some of them (51A, 51B, 51D, and 51H) could be identified by immunological methods, and it was also shown by single-cell immunofluorescence that these proteins are indeed present on the surfaces of individual cells (26). With the new knowledge about the genetic properties of the entire *SAg* family, it was now possible to characterize the *Rdr3* knockdown phenotype of all *SAg* genes. To determine which of the genes identified here is controlled by this RNAi pathway, we analyzed wild-type-expressing (serotype 51A) cultures and *Rdr3* knockdown cultures by cDNA microarrays. The following analysis combines the expression data from the microarray with the chromosomal localization of genes and distinguishes between the distance of genes to the closest telomeric site (thus including the data on chromosome heterogeneity) and the distance to the end of scaffolds (thus ignoring chromosome heterogeneity), as illustrated in Fig. 6A.

Figure 6B shows the intensity of *SAg* genes in wild-type cultures in reference to the ends of scaffolds, and Fig. 6C shows the intensity in reference to the closest telomeric region. Both figures show one dominantly expressed *SAg* (*51/29 A*) and only four other genes (*SAg 5*, *3*, *56*, and *2*) with intensities above background; however, the values were ~5 times lower than that of the activated 51A gene. SAg 3 is divergent from all other SAgs in the phylogenetic analysis; *SAg 5* and *56* are ohnologs from last WGD, and *SAg 2* shows an ohnolog of the intermediate WGD. All other genes do not show signals above threshold. Therefore, we conclude that the majority of newly identified *SAg* genes (ohnologs and nonohnologs) are not expressed during vegetative growth.

Figure 6D and E show the situation in *Rdr3* knockdown cultures in reference to the wild type (the fold change relative to wild-type levels is shown in Fig. 6B and D). Again, Fig. 6D shows the genes in reference to the end of scaffolds, and Fig. 6E shows them in reference to the closest telomere. Both figures do not show a general coexpression of all genes but show only 14 genes above background level. As Fig. 6D shows upregulated genes with no clear pattern, this can be identified in Fig. 6E, which includes the information on chromosome heterogeneity, indicating that only genes with a short distance to telomeric repeats are upregulated in *Rdr3* knockdown cells. Within this group of genes, the maximum distance to a predicted telomeric site is ~20 kb. Note that the *51A* gene was also still highly expressed in these cultures but did not show any differential change in the microarray. Figure 6F shows the upregulated genes with decreasing levels: these are mostly represented by classical surface antigens and isoforms and two non-annotated genes, *SAg 27* and *SAg 32*.

However, silencing of *SAg* genes apparently involves two distinct mechanisms, as Rdr3-mediated silencing occurs exclusively in subtelomeric regions. *SAg* activation induced by *Rdr3* knockdown also occurs only at subtelomeric loci. This is the first demonstration of RNAi-mediated subtelomeric silencing; this phenomenon is restricted to the surface antigens, as we do not see a general activation of subtelomeric genes in *Rdr3* silencing: Fig. 6G shows the remaining non-*SAg* genes which show significant upregulation in reference to the closest telomeric site. No preference for subtelomeric genes can be identified in Fig. 6G.

Linking telomere position effect and homology-dependent silencing. Gene position cannot be the only parameter controlling silencing of mutually exclusive expressed *SAg* genes, because several *SAg* genes with close proximity to telomeric repeats are not upregulated during *Rdr3* knockdown (Fig. 6E). Also, we did not see a genome-wide upregulation of subtelomeric genes. As only particular subtelomeric genes are under the control of Rdr3, a general spreading of the heterochromatic state of telomeres cannot be responsible.

Homology dependency of serotype expression was shown in a series of previous studies (reviewed in reference 19). As serotype expression requires a genetic cross talk between silent and active genes, the knowledge that RNAi controls serotype expression suggests that the *trans*-acting factor could be short RNAs. Further support comes from experiments in which the actively transcribed serotype gene is silenced: introduction of siRNAs against the active gene not only knocked down its expression in *cis* but also activated formerly silent *SAg* genes in *trans* (41). Although their function is still not clear, endogenous small RNAs of silent as well as active *SAg* genes can be demonstrated in Northern blots, thus indicating that a complex genetic cross talk is involved in serotype regulation (our unpublished observations). We extended the analysis to predict possible *trans*-acting siRNAs from the sequence data. Figure 7 shows a heat map of possible 20-nucleotide (nt) siRNAs, indicating a high degree of identity between individual *SAg* genes, which becomes apparent by the high numbers of po-

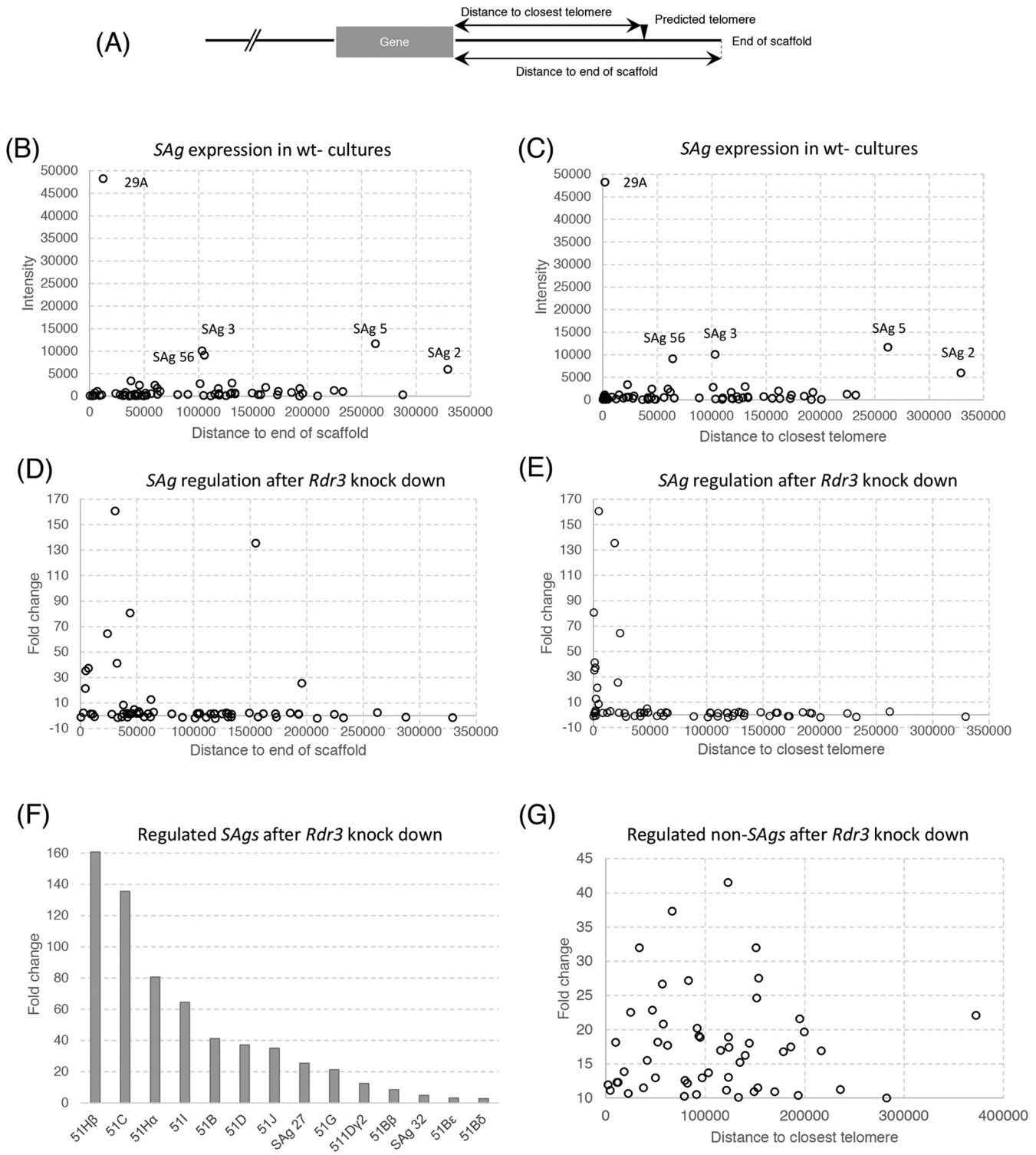

FIG 6 Microarray analysis of *SAg* expression in wild-type and Rdr3 knockdown cultures in relation to chromosomal localization. The graphs show expression levels as functions of intensity in reference to the distance to (i) the end of scaffold, ignoring macronuclear heterogeneity, and (ii) the closest telomeric site, thus also taking the macronuclear heterogeneity into account (illustrated in panel A). (B) Intensity of microarray analysis of a pure serotype 51A culture in relation to gene distance to the end of assembled scaffolds. (C) Intensity of microarray analysis of a pure serotype 51A expressing culture in relation to gene distance to the closest telomeric site. Genes with expression levels significantly above background are indicated. (D) Fold change relative to the wild-type expression level after 9 days of silencing of Rdr3 in relation to gene distance to the ends of scaffolds. (E) Fold change relative to wild-type expression levels after 9 days of silencing of Rdr3 in relation to the closest telomeric site. (F) Significantly upregulated genes (fold change $> +2.6$) with decreasing levels of upregulation. (G) Gene positions of the remaining non-*SAg* genes which show upregulation in Rdr3 knockdown cultures in relation to the closest telomeric site.

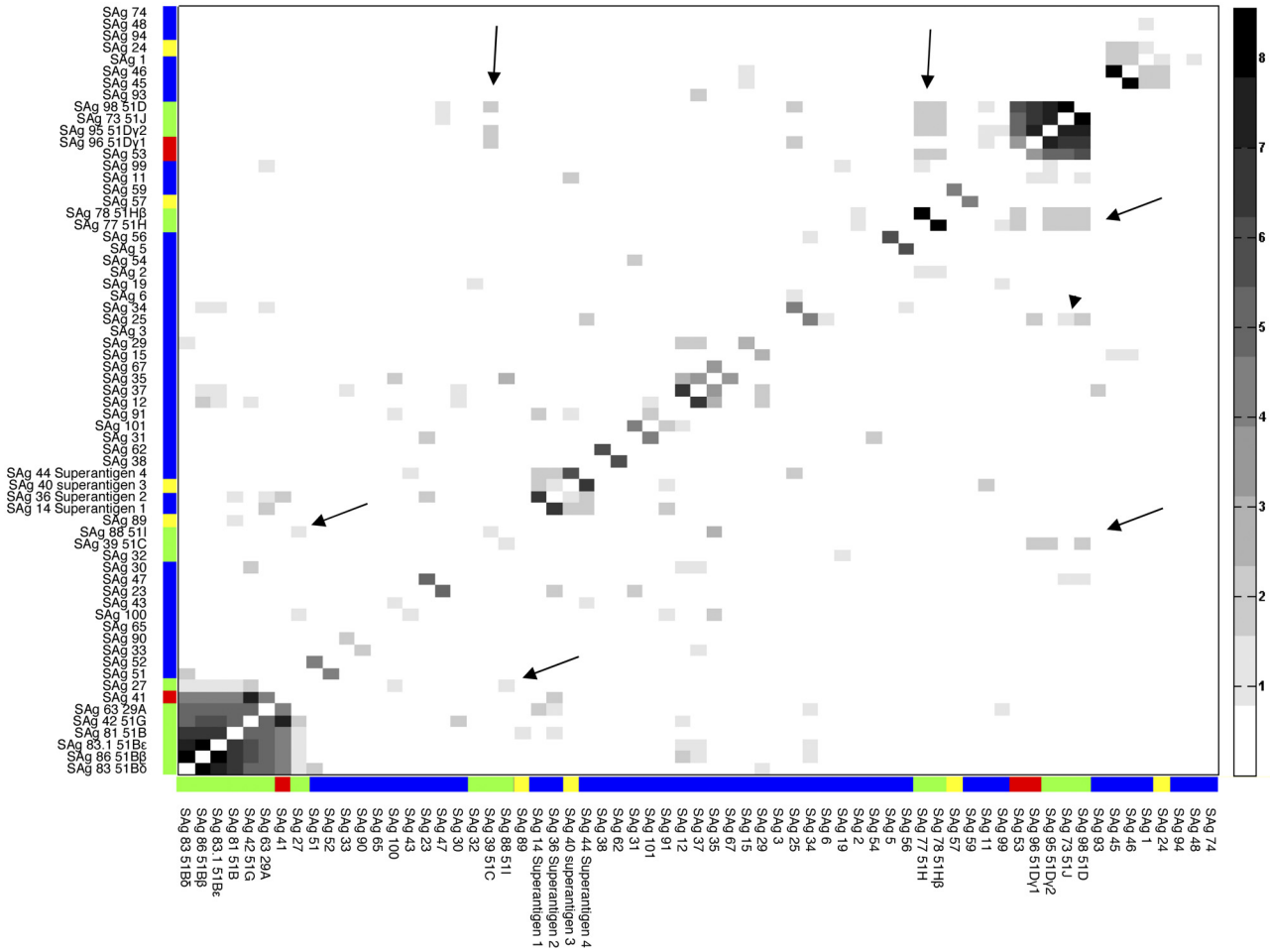

FIG 7 Heat map of potential *trans*-acting siRNAs between CDS of *SAg* genes. Gray squares indicate the number of perfect matching 20-nt stretches. The legend on the right indicates natural logarithmic numbers of mutual 20nt-meres + 1. Upregulated genes in Rdr3 knockdown cultures are indicated with green. Unregulated genes of an individual cluster with a greater distance to telomeric repeats are indicated with red. Subtelomeric genes which do not show upregulation in Rdr3 knockdown cultures, are indicated with yellow, and intrachromosomal *SAg* genes are indicated with blue. All genes are listed in the order determined by the phylogenetic analysis. Arrows indicate clusters of potential siRNAs linking the D cluster with antigens 51Hα + 51Hβ and 51C, and the large cluster consisting of the *51A*, *51B*, and *51G* genes with the *51I* gene. Note the asymmetry of the map as, e.g., the *51J* gene has more siRNAs in common with *SAg 25* than vice versa (arrowhead) because of internal repeat structures.

tential *trans*-acting siRNAs. This length was chosen because ~20-nt siRNAs were observed in Northern blots probed with *SAg*-specific probes. For the heat map, we used the order of genes resulting from the phylogenetic analysis in Fig. 1A. The heat map indicates two large clusters of potential *trans*-acting siRNA producing genes, one consisting of the *51A-51B-51G* group and one consisting of the *51D-51J* group. The latter also shows a large amount of potential siRNAs specific to both *51H* genes and the *51C* gene (arrows), although these are clearly separated in the phylogenetic tree (Fig. 1A).

We combined siRNA analysis with the expression data, and the Rdr3-controlled *SAg* genes are indicated in Fig. 7 by green boxes. Their pattern is in keeping with the above-mentioned clusters of potential *trans*-acting siRNA-producing genes, as *Rdr3* knockdown-mediated *SAg* activation occurs exclusively at loci which share *trans*-acting siRNAs with others. Exceptions are *SAg 41* in the first cluster and *SAg 53* and *51Dγ-1* in the second large cluster (Fig. 7; red boxes). In agreement with our conclusion that subtelomeric gene position is a prerequisite for the underlying expression mechanism, both genes show the individual largest distance to telomeric repeats inside the cluster (*SAg 41*, ~42 kb; *SAg 53*, ~148 kb; *51Dγ-1*, ~12.5 kb). The only upregulated gene for which Fig. 7 does not give a satisfying explanation is *SAg 32*, because it does not show possible *trans*-acting siRNAs specific to any cluster.

The theoretical calculation of potential *trans*-acting siRNAs provides a potential explanation as to why these *SAg* genes in particular are controlled by the Rdr3-related RNAi mechanism and supports the hypothesis that *trans*-acting siRNAs are involved in serotype regulation, because the Rdr3-regulated *SAg* genes require subtelomeric localization and *trans*-acting siRNAs. In agreement with our hypothesis, all other *SAg* genes, even the subset of genes which also show close proximity to telomeric repeats (*SAg* genes 24, 89, 57, 99, 40, and 101) (Fig. 7; yellow boxes), show only a limited overlap with other *SAg* genes (Fig. 7) and do not show significant upregulation. In conclusion, our data indicate that Rdr3-mediated silencing occurs only in subtelomeric genes which share *trans*-acting siRNAs with other subtelomeric genes.

DISCUSSION

The *SAg* multigene family. In order to understand the mechanisms of mutually exclusive expression of different *SAg* genes in *P. tetraurelia*, it was necessary to characterize all the *SAg* genes present in the genome. Sixty-five SAgs had the characteristic cysteine periodicity and included 8 classical serotype genes (alphas) and 6 isogenes. However, 8 potential SAgs did not possess a GPI anchor and instead had transmembrane domains. It is not known what the function of these genes is, but as five of them do not show an ER-TLS, their function may be intracellular. Sixty-four *SAg* genes seemed to have structures compatible with expression (i.e., they were probably not pseudogenes). The codon usage with high CAI values was suggestive of highly expressed genes. Although most *SAg* genes are silent at some time, this is also consistent with the alpha *SAg* genes and their isoforms; however, those in particular have a codon bias which seems optimized for high-level expression. In fact, this correlates with our finding that only those genes can be activated during Rdr3 silencing as well as in serotype pure wild-type cultures, and it also agrees with reports that serotype proteins contribute to ~3.5% of the total cellular proteins (42). With respect to the large open reading frames (ORFs), an efficient optimization of transcription and translation is indeed required to ensure high-level protein expression. SAgs are high-molecular-weight proteins encoded by very long genes. Introns are definitely underrepresented, which may be due to optimization of transcription, although the tiny introns of *Paramecium*, which have a size between 20 and 34 nt (43), would not create such a drastic increase in transcript size.

In some cases, in addition to the major alpha gene, there are neighboring isogenes in the chromosome, and reverse transcription-PCR (RT-PCR) experiments (Fig. 5) showed that some of these were also transcribed. An attractive explanation for intrachromosomal coexpression in wild-type cells comes from studies in *Plasmodium* indicating a nuclear organization which allows the subnuclear translocation of loci into euchromatic regions, which allow transcriptional activity (44). A similar locus repositioning in *Paramecium* macronuclei might explain cotranscription of close loci (isogenes) on the same macronuclear chromosomes, whereas other loci showing a comparable degree of identity and/or *trans*-acting siRNAs which are located on different chromosomes are not coactivated. In support of this hypothesis, *in situ* localization of *SAg* transcripts indicated local transcription spots in the macronucleus (45).

A telomere position effect controls RNAi-dependent gene regulation. Although, in *Paramecium*, only one serotype is usually expressed, knockdown of Rdr3 results in expression of a mixture of SAgs (26). In this paper, we show that only *SAg* genes with a subtelomeric location show this Rdr3-dependent silencing (Fig. 6).

One possibility to explain subtelomeric position of the genes is the involvement of the telomeric heterochromatin state in the expression mechanism. Telomeric and subtelomeric chromatin is usually heterochromatic. It is likely that Rdr3-associated siRNAs are involved in transcriptional silencing in *Paramecium* (26), and the coding region has been shown to control transcriptional activity of *SAg* genes (46). A precise understanding of the mechanisms involved will require characterization of the siRNAs present during the mutually exclusive expression of *SAg* genes as well as the changes occurring during phenotypic variation. However, Fig. 7 shows that there is a relationship between mutually exclusive expression and the occurrence of potential *trans*-acting siRNAs, thus suggesting that such siRNAs are involved in the genetic cross talk enabling mutually exclusive expression. The details of siRNA source and targets need to be characterized at the molecular level to get an impression of the RNA-mediated networks occurring in the nucleus.

In general, endogenous *trans*-acting siRNAs have great importance for the regulation of endogenous gene expression, as they allow genetic cross talk and therefore flexible adaptation of a single gene as well as alterations in gene expression patterns. Position effects in correlation with *trans*-acting mechanisms have been reported for *Schizosaccharomyces pombe*, where silencing depends on the position of the target gene: *trans*-silencing works efficient only in loci close to heterochromatic regions, such as centromeres (47). A different gene position effect was shown in *Drosophila*: transgenes inserted in subtelomeric regions repress expression of homologous genes in euchromatin, and this *trans*-silencing has also been shown to depend on components of the piRNA (piwi-interacting RNA) pathway, which is a germline-specific RNAi mechanism, and on heterochromatin protein 1 (HP1); interestingly, epigenetic inheritance has also been shown, and increasing data indicate cytoplasmically transmitted piRNAs responsible for maternal inheritance (48, 49). Although the two examples cannot be directly compared to each other, increasing evidence shows gene position effects as prerequisites for efficient *trans*-silencing; our example of *SAg* regulation in *Paramecium* would be one of the first examples in regulation of endogenous gene expression. Reports that these *trans*-silencing phenomena also allow epigenetic inheritance, as has also been reported for the *Paramecium* serotype system (10, 28), could give rise to speculations that the telomere position effect may also be necessary for stable inheritance of gene expression patterns.

Diversification of serotypes by gene duplication and chromosome fragmentation. Epigenetic control of antigenic variation represents an exciting example combining Darwinian evolution of surface antigen genes as a kind of long-term adaptation with the capability for rapid phenotype alterations by epigenetic control of a variety of these genes and ciliate genetics and epigenetics heavily contributes to a better understanding of these rarely understood control mechanisms (10). Across kingdoms, phenotypic evolution was described as resulting from the creation of new genes, predominantly originating from duplication events (50). The example of variant surface antigens has another aspect: as the creation of a new functional surface antigen also requires consideration of the gene expression mechanism (to include the newly evolved gene into the genetic network of telomeric *trans* silencing), translocation of the gene copy into a subtelomeric region is required.

Our analysis of the *SAg* family in *Paramecium* indicates its evolution by two different events of gene duplication: a subset of intrachromosomal genes show ohnologs from the *Paramecium* intermediate and recent WGD, and another subset of subtelomeric genes show intrachromosomal gene duplicates which need to result from a different mechanism. Strikingly, the latter subset of genes show (i) mutually exclusive transcription, (ii) intrachromosomal duplicates, and (iii) no ohnologs. As 68% of all genes in *P. tetraurelia* still have their ohnologs from the last WGD, we see only 25 ohnologs in the *SAg* family, which has 65 genes.

Why do the subtelomeric *SAg* genes not have ohnologs? They

might be eliminated soon after the WGD event, or subtelomeric *SAg* genes might have appeared after the last WGD. Both possibilities may also occur together. A preferential deletion of subtelomeric *SAg* ohnologs might be selected due to problems in the mechanism for mutually exclusive expression. Our data indicate that closely related *SAg* genes on different chromosomes are usually mutually exclusively expressed, e.g., *51A-51G* or *51D-51J*, by the above-discussed *trans*-acting siRNAs. Previous studies showed allelic exclusion in cells which are heterozygous for *SAg* loci, meaning that one of the two alleles became inactivated: also, several crosses and backcrosses led to the hypothesis that this silencing was due to an "allelic interaction" (51). Allelic exclusion of very similar subtelomeric alleles or *trans* silencing of duplicates might explain a low retention rate of subtelomeric *SAg* ohnologs.

Nevertheless, our data suggest gene duplication events after the last WGD contributing to *SAg* diversity. The phylogenetic analysis indicates a closer relationship between intrachromosomal duplicates than ohnologs, indicating that intrachromosomal isoforms are younger than the last WGD, which is in agreement with the absence of ohnologs. As they show cotranscription at a certain level with the individual alpha gene, this might explain the still existing selection pressure on (i) cysteine periodicity and (ii) N- and C-terminal areas. As a consequence of cotranscription, the minor abundance of isoforms together with the alpha serotype protein might allow diversification of the central and variable area of the proteins with parallel positive selection of tertiary structure.

The finding of intrachromosomal gene duplicates on macronuclear chromosomes raises the question of their origin. As one would expect from highly expressed intronless genes, retroposition might meet several criteria of these isogenes, e.g., lack of introns and high conservation of coding sequences (CDS) but not of regulatory up- and downstream regions. However, retroposition can be excluded, as all isogenes are also present in the micronuclear genome and all of them contain IES elements, which themselves show a high degree of similarity (data not shown). In ciliates, the germline chromosomes are interrupted by noncoding internal-elimination elements (IESs) which are present in coding and noncoding regions: their precise elimination during assembly of the new somatic macronucleus is crucial for building intact genes (52). The data of a recent sequencing of *Paramecium* IESs (53) also show IESs in the isogenes which show great homology to the IESs of the individual alpha gene, thus indicating that duplication occurs on micronuclear chromosomes.

What might be the biological significance of these intrachromosomal duplicates? Mutually exclusive transcription requires subtelomeric localizations; therefore, the evolution from intrachromsomal *SAg* isoforms into a new *SAg* would require gene repositioning. The observed gene position effects in context with macronuclear heterogeneity let us hypothesize that serotype evolution in ciliates involves a very elegant solution, as the heterogeneity of macronuclear chromosomes allows them to "move the telomere" instead of the gene. The appearance of the telomeric sites close to the *SAg* isoforms suggests that alterations of macronuclear chromosome heterogeneity by the creation/activation of new telomere addition sites on micronuclear chromosomes may be a mechanism to advance a *SAg* isoform to a new serotype gene by moving the gene into a subtelomeric region (see the model in Fig. 8). As mentioned above, previous studies showed that imprecise elimination of repeated sequences (transposable elements or minisatellites) from micronuclear chromosomes leads to macronuclear chromosome polymorphisms: after deletion, the ends may be stabilized by telomere addition or religated (54, 55). These imprecise deletions are epigenetically controlled: they can be experimentally induced by prezygotic silencing (56), and a non-Mendelian mutant was described which still contains the *51A* gene in the micronucleus but removes this locus during macronuclear development, showing an alternative telomere addition site upstream of the *51A* gene (39). Although the detailed mechanisms remain to be clarified, these findings indicate that epigenetic mechanisms can contribute to the creation of macronuclear heterogeneity, which could move internal *SAg* duplicates into subtelomeric regions.

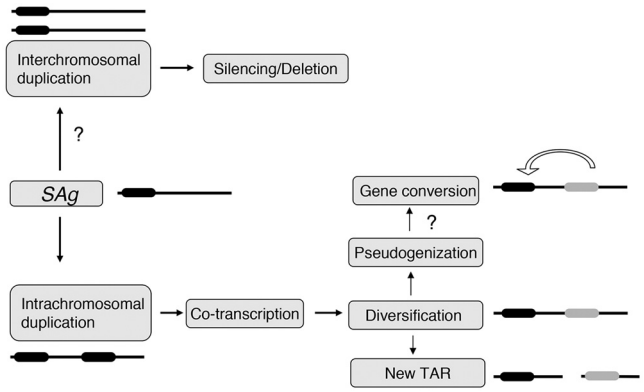

FIG 8 Hypothetical model of serotype gene evolution. Gene duplication occurs in the micronucleus. As subtelomeric *SAg* genes do not show ohnologs, this may indicate that duplicates from WGD become deleted or that all subtelomeric *SAg* genes are younger than the last WGD and therefore cannot show ohnologs. In contrast, gene duplications leading to *SAg* genes on the same Mac chromosome show cotranscription in a homology-dependent manner, allowing diversification of the intrachromosomal duplicate with selection pressure on structurally relevant areas. This may lead to new immunological information on the duplicate with decreasing levels of cotranscription as a result of decreased homology. Genetic and epigenetic alterations driving alternative TAR (telomere addition regions) close to the intrachromosomal duplicate can move the divergent duplicate close to a telomere, advancing this gene into a new serotype gene, as the new subtelomeric localization allows mutually exclusive expression. Alternatively, the intrachromosomal duplicate may become a pseudogene, allowing rapid diversification by a loss of selection pressure, and serves as a source of new immunological information by copying the information on the subtelomeric serotype gene by gene conversion.

As gene conversion was shown to be involved in the creation of new variant antigen genes in many organisms, such as *Trypanosoma* spp. (57), cells may benefit from the intrachromosomal duplicates in different ways, because they offer a pool of new immunological information: as studies indicate the frequency of gene conversion to be inversely proportional to physical distance (58), this may be a mechanism to introduce new immunological data into the existing subtelomeric *SAg*.

MATERIALS AND METHODS

Cell cultures and RNAi. *Paramecium tetraurelia* strains d4-2 and 51 were used and cultured as described before in WGP (wheat-grass powder) medium inoculated with *Klebsiella pneumoniae* and supplemented with 0.8 μg/ml β-sitosterol. Serotype-pure cultures were cultivated at 31°C (51A), 26°C (51D), 22°C (51B), and 12°C (51H). RNAi by feeding against *Rdr3* and *Icl7a* (control) genes was carried out as described in references 26 and 59: fragments of the coding region were cloned into the plasmid L4440 and transformed into the RNase III-deficient *Escherichia coli* strain HT115DE3. After growth of bacteria in LB medium and induction of

double-stranded-RNA (dsRNA) synthesis by addition of IPTG, dsRNA synthesis was controlled in an aliquot by extraction by acid-phenol. The positions of the fragments used were 1789 to 2462 for *Rdr3* (GSPATT00006401001) and 1 to 580 for *Icl7a* (GSPATG00021610001). For serotype analysis, ~50 cells were incubated with 1/100 homologous antiserum (anti-51A, anti-51B, anti-51D, and anti-51H; kind gift of James D. Forney, Purdue University, West Lafayette, IN, USA), and immobilization was quantified after 10 min.

Proteome analysis, gene characteristics, and phylogenetic analysis. *SAg* genes were identified by a proteome-wide search for against the Pfam library of hidden Markov models (HMMs) (60) and extraction of hits for the *Paramecium* surface antigen domain Pfam *Paramecium_SA* (PF01508), which is based on eight-cysteine periods, as described for the *Paramecium primaurelia* 156G protein (61). We used a threshold of an *E* value of $\leq 10^{-9}$. The resulting genes were manually reannotated from the original Sanger reads of the *Paramecium tetraurelia* genome project (36). Introns were verified by RNAseq data. Sequence data have been deposited at ParameciumDB and integrated in the new gene annotation (O. Arnaiz et al., unpublished data); accession numbers can be found in Table S1. SAg sequences are also included in the supplemental material.

Internal tandem repeats were identified with the tandem repeat finder tool using a score of ≥319 (62). N-terminal ER translocation signals (ER-TLS) were predicted by the SignalIP 3.0 software (63), and C-terminal signal peptides for GPI anchoring were predicted using Kohonen self-organizing map GPI-SOM (64). Transmembrane domains were predicted by TMHMM as described in reference 65. Graphic alignments were calculated by GATA using default settings (66). Amino acid sequences of SAgs were aligned with ClustalW, and the neighbor joining tree was calculated with MEGA4 (67) with 1,000 bootstrap replicates, based on *p* distances and after pairwise deletion of gaps.

Northern blotting and real-time PCR. Total RNA from vegetative cultures was isolated with Trizol (Invitrogen, Karlsruhe, Germany) according to the manufacturer's instructions. For Northern blots, 10 μg of total RNA was separated on a 1.2% formaldehyde agarose gel and capillary blotted in 10× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) to Hybond N+ membranes (GE, Braunschweig, Germany) and UV cross-linked. Hybridizations to mRNA and rRNA were carried out at 65° in 1× Church buffer (7% SDS, 0.25 M sodium phosphate, 1% SDS, 1 mM EDTA [pH 7.2]). Membranes were washed in 2× SSC, 0.1% SDS and subsequently in 0.2× SSC, 0.1% SDS before exposure to phosphorimager plates. Probes were produced by random priming with [α-^{32}P]dCTP (3,000 Ci/mmol).

For analysis of transcript levels by real-time RT-PCR, total RNA was isolated with Trizol (Invitrogen, Karlsruhe, Germany) according to the manufacturer's instructions and additionally purified with the RNeasy microkit (Qiagen, Hilden, Germany). RNA (500 ng) was then reverse transcribed with Moloney murine leukemia virus (M-MuLV) reverse transcriptase (NEB, Frankfurt am Main, Germany). Quantitative PCR was carried out with the EvaGreen qPCR mix (Axon, Kaiserslautern, Germany) on a CFX96 detection system (Bio-Rad, Hercules, CA). Relative quantification was calculated (as described in reference 68) in reference to expression to GAPDH, which is constitutively expressed in *Paramecium* (41). See Table S2 in the supplemental material for primer sequences and positions as well as for probe positions.

Microarray analysis. *Rdr3* and *Icl* RNAi cultures were fed with dsRNA-producing bacteria for 9 days with three biological replicates each. Total RNA was isolated with the RNAspin II kit (Machery & Nagel, Düren, Germany). After additional digestion with DNase and subsequent purification with acid phenol, integrity was checked by analysis on the Bioanalyzer 2100 RNA 6000 nanochip (Agilent, Böblingen, Germany). Microarray analysis was carried out at PartnerChip, Évry, France: 300 ng total RNA was reverse transcribed using the TransPlex whole-transcriptome amplification kit (Sigma-Aldrich, Seelze, Germany) according to the manufacturer's protocol, and the library was subsequently amplified by 15 cycles with Platinum *Taq* polymerase (Invitrogen, Karlsruhe, Germany). The cDNA was then purified with the GenElute cleanup kit (Sigma, Seelze, Germany), and 1 μg of cDNA was labeled with Cy3 using Klenow fragment. Labeled cDNA (12 μg) was used for hybridization by incubation in the NimbleGen hybridization system for 24 h using a custom expression array from the *Paramecium* genome (3 × 720K array format; RocheNimbleGen Inc., Madison, WI). The array design consists of one set (SET01) of probes for annotated genes described in reference 69 and a redesigned set (SET02) taking improved gene models into account and using 12 probes per gene to discriminate between highly similar genes. Arrays were washed three times before being dried and scanned. Raw data were extracted using the NimbleScan software and analyzed by GeneSpring GX11 (Agilent, Böblingen, Germany). The signal-to-noise ratios were higher than 3. For each transcript, the signal was calculated after RMA (robust multiarray average) normalization according to reference 70. Pearson correlation coefficients between biological samples were high, ranging from 0.907 to 0.997, indicating good technical replication. Microarray data have been deposited at the Gene Expression Omnibus database (71) under accession number GSE59390.

Identification of telomeric sites. A Perl script was used to extract the Sanger reads from the *Paramecium* genome project (36) containing at least three telomeric repeats (CCC[CA]AA) with no more than one mismatch. The mapping to the reference genome was performed using local mode of the Bowtie 2 aligner with default settings (72). A Perl script was written to estimate the position of telomeric sites in the genome. To distinguish singletons from a relevant signal, a criterion of at least 4 overlapping reads in the range of 2 kb was used. Because only the genomic part of the read was mapped, the alignment position of the right endpoint of the read was taken as the telomere position. Finally, the telomeric site was calculated as the median of the individual read positions which met the above-mentioned requirements.

Calculation of the dN/dS ratio and the codon adaptation index. MATLAB scripts were written to calculate the dN/dS ratio and the codon adaptation index (MATLAB and Statistics Toolbox, release 2012b; The MathWorks, Inc., Natick, MA, USA). The dN/dS ratio was calculated using the Nei-Gojobori method (http://www.mathworks.com/help/bioinfo/ref/dnds.html) and the sliding window with window size ranging from 450 to 1,000 bp. The CAI was calculated using the formula from reference 73. Highly expressed actin CDS (74) were used as a reference to calculate the relative codon adaptiveness table.

Data accession numbers. *SAg* sequence data have been deposited at ParameciumDB and integrated in the new gene annotation (O. Arnaiz et al., unpublished data); accession numbers can be found in Table S1. Microarray data have been deposited at the Gene Expression Omnibus database (71) under accession number GSE59390.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.01328-14/-/DCSupplemental.

Figure S1, PDF file, 2 MB.
Figure S2, PDF file, 0.1 MB.
Figure S3, PDF file, 0.1 MB.
Figure S4, PDF file, 0.1 MB.
Table S1, PDF file, 1.5 MB.
Table S2, PDF file, 0.8 MB.
Text S1, TXT file, 0.5 MB.

ACKNOWLEDGMENTS

This work was supported by DFG grant SI-1397-2 to M.S. D.B. received a scholarship of the DAAD (A/12/92126). M.C. received an excellence graduate fellowship from the University of Kaiserslautern.

This study and in particular the microarray design were performed in the context of the CNRS-supported European Research Group GDRE "*Paramecium* Genome Dynamics and Evolution." We are grateful to all members of the GDRE and to the European Science Foundation COST network BM1102 "Ciliates as model systems to study genome evolution, mechanisms of non-Mendelian inheritance, and their roles in environ-

mental adaptation." We thank Angelika Preisfeld, University of Wuppertal, for critical comments on the manuscript, Eric Meyer, Ecole Normale Superieur, Paris, for sharing ideas and discussion, and Linda Sperling and Olivier Arnaiz, Gif-sur-Yvette, for kind support.

REFERENCES

1. **Borst P.** 2003. Mechanisms of antigenic variation: an overview, p 1–15. *In* Craig A, Scherf A (ed), Antigenic variation. Elsevier, Oxford, United Kingdom.
2. **Rössle R.** 1905. Spezifische Seren gegen Infusorien. Arch. Hyg. Bakteriol. **54:**1–31.
3. **Clark TG, Forney JD.** 2003. Free-living and parasitic ciliates, p 3751–402. *In* Craig A, Scherf A (ed), Antigenic variation. Elsevier, Oxford, United Kingdom.
4. **Simon MC, Schmidt HJ.** 2007. Antigenic variation in ciliates: antigen structure, function, expression. J. Eukaryot. Microbiol. **54:**1–7. http://dx.doi.org/10.1111/j.1550-7408.2006.00226.x.
5. **Harumoto T, Miyake A.** 1993. Possible participation of surface antigens of *Paramecium* in predator-prey interaction. J. Eukaryot. Microbiol. **40:**27A.
6. **Harumoto T.** 1994. The role of trichocyst discharge and backward swimming in escaping behaviour of *Paramecium* from *Dileptus margaritifer*. J. Eukaryot. Microbiol. **41:**560–564. http://dx.doi.org/10.1111/j.1550-7408.1994.tb01517.x.
7. **Simon MC, Kusch J.** 2013. Communicative functions of GPI-anchored surface proteins in unicellular eukaryotes. Crit. Rev. Microbiol. **39:**70–78. http://dx.doi.org/10.3109/1040841X.2012.691459.
8. **Deitsch KW, Lukehart SA, Stringer JR.** 2009. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. Nat. Rev. Microbiol. **7:**493–503. http://dx.doi.org/10.1038/nrmicro2145.
9. **Wyse BA, Oshidari R, Jeffery DC, Yankulov KY.** 2013. Parasite epigenetics and immune evasion: lessons from budding yeast. Epigenetics Chromatin **6:**40. http://dx.doi.org/10.1186/1756-8935-6-40.
10. **Simon M, Plattner H.** 2014. Unicellular eukaryotes as models in cell and molecular biology: critical appraisal of their past and future value. Int. Rev. Cell Mol. Biol. **309:**141–198. http://dx.doi.org/10.1016/B978-0-12-800255-1.00003-X.
11. **Taylor JE, Rudenko G.** 2006. Switching trypanosome coats: what's in the wardrobe? Trends Genet. **22:**614–620. http://dx.doi.org/10.1016/j.tig.2006.08.003.
12. **Navarro M, Peñate X, Landeira D.** 2007. Nuclear architecture underlying gene expression in *Trypanosoma brucei*. Trends Microbiol. **15:**263–270. http://dx.doi.org/10.1016/j.tim.2007.04.004.
13. **Scherf A, Hernandez-Rivas R, Buffet P, Bottius E, Benatar C, Pouvelle B, Gysin J, Lanzer M.** 1998. Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in *Plasmodium falciparum*. EMBO J. **17:**5418–5426. http://dx.doi.org/10.1093/emboj/17.18.5418.
14. **Prucca CG, Slavin I, Quiroga R, Elías EV, Rivero FD, Saura A, Carranza PG, Luján HD.** 2008. Antigenic variation in *Giardia lamblia* is regulated by RNA interference. Nature **456:**750–754. http://dx.doi.org/10.1038/nature07585.
15. **Merrick CJ, Duraisingh MT.** 2006. Heterochromatin-mediated control of virulence gene expression. Mol. Microbiol. **62:**612–620. http://dx.doi.org/10.1111/j.1365-2958.2006.05397.x.
16. **Hernandez-Rivas R, Pérez-Toledo K, Herrera Solorio AM, Delgadillo DM, Vargas M.** 2010. Telomeric heterochromatin in *Plasmodium falciparum*. J. Biomed. Biotechnol. 2010:290501. PubMed. http://dx.doi.org/10.1155/2010/290501.
17. **Ralph SA, Scherf A.** 2005. The epigenetic control of antigenic variation in *Plasmodium falciparum*. Curr. Opin. Microbiol. **8:**434–440. http://dx.doi.org/10.1016/j.mib.2005.06.007.
18. **Sonneborn TM, Lesuer A.** 1948. Antigenic characters in *Paramecium aureliae* (variety-4)—determination, inheritance and induced mutations. Am. Nat. **82:**69–78. http://dx.doi.org/10.1086/281566.
19. **Prat A.** 1990. Conserved sequences flank variable tandem repeats in two alleles of the G surface protein of Paramecium primaurelia. J. Mol. Biol. **211:**521–535. http://dx.doi.org/10.1016/0022-2836(90)90263-L.
20. **Nielsen E, You Y, Forney J.** 1991. Cysteine residue periodicity is a conserved structural feature of variable surface proteins from *Paramecium tetraurelia*. J. Mol. Biol. **222:**835–841. http://dx.doi.org/10.1016/0022-2836(91)90573-O.
21. **Carrington M, Miller N, Blum M, Roditi I, Wiley D, Turner M.** 1991. Variant specific glycoprotein of *Trypanosoma brucei* consists of two domains each having an independently conserved pattern of cysteine residues. J. Mol. Biol. **221:**823–835. http://dx.doi.org/10.1016/0022-2836(91)80178-W.
22. **Ferguson MA.** 1999. The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research. J. Cell Sci. **112:**2799–2809.
23. **Capdeville Y, Benwakrim A.** 1996. The major ciliary membrane proteins in *Paramecium primaurelia* are all glycosylphosphatidylinositol anchored proteins. Eur. J. Cell Biol. **70:**339–346.
24. **Simon MC, Schmidt HJ.** 2005. Variety of serotypes of *Paramecium primaurelia*: single epitopes are responsible for immunological differentiation. J. Eukaryot. Microbiol. **52:**319–327. http://dx.doi.org/10.1111/j.1550-7408.2005.00040.x.
25. **Capdeville Y, Caron F, Antony C, Deregnaucourt C, Keller AM.** 1987. Allelic antigen and membrane anchor epitopes of *Paramecium primaurelia* surface antigens. J. Cell Sci. **88:**553–562.
26. **Marker S, Le Mouël A, Meyer E, Simon M.** 2010. 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. http://dx.doi.org/10.1093/nar/gkq131.
27. **Chalker DL, Stover NA.** 2007. Genome evolution: a double take for *Paramecium*. Curr. Biol. **17:**R97–R99. http://dx.doi.org/10.1016/j.cub.2006.12.002.
28. **Preer JR, Preer LB, Rudman BM, Barnett AJ.** 1985. Deviation from the universal code shown by the gene for surface protein 51A in *Paramecium*. Nature **314:**188–190. http://dx.doi.org/10.1038/314188a0.
29. **Scott J, Leeck C, Forney J.** 1993. Molecular and genetic analyses of the B type surface protein gene from *Paramecium tetraurelia*. Genetics **134:**189–198.
30. **Breuer M, Schulte G, Schwegmann KJ, Schmidt HJ.** 1996. Molecular characterization of the D surface protein gene subfamily in *Paramecium tetraurelia*. J. Eukaryot. Microbiol. **43:**314–322. http://dx.doi.org/10.1111/j.1550-7408.1996.tb03994.x.
31. **Duharcourt S, Keller AM, Meyer E.** 1998. Homology-dependent maternal inhibition of developmental excision of internal eliminated sequences in *Paramecium tetraurelia*. Mol. Cell. Biol. **18:**7075–7085.
32. **Godiska R.** 1987. Structure and sequence of the H surface protein gene of *Paramecium* and comparison with related genes. Mol. Gen. Genet. **208:**529–536. http://dx.doi.org/10.1007/BF00328151.
33. **Arnaiz O, Sperling L.** 2011. ParameciumDB in 2011: new tools and new data for functional and comparative genomics of the model ciliate *Paramecium tetraurelia*. Nucleic Acids Res. **39:**D632–D636. http://dx.doi.org/10.1093/nar/gkq918.
34. **Preer JR.** 1959. Studies on the immobilization antigens of *Paramecium*. IV. Properties of the different antigens. Genetics **44:**803–814.
35. **Schmidt HJ.** 1987. Characterization and comparison of genomic DNA clones containing complementary sequences to mRNA from serotype 51D of *Paramecium* tetraurelia. Mol. Gen. Genet. **208:**450–456. http://dx.doi.org/10.1007/BF00328138.
36. **Aury JM, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Ségurens B, Daubin V, Anthouard V, Aiach N, Arnaiz O, Billaut A, Beisson J, Blanc I, Bouhouche K, Câmara F, Duharcourt S, Guigo R, Gogendeau D, Katinka M, Keller AM, Kissmehl R, Klotz C, Koll F, Le Mouël A, Lepère G, Malinsky S, Nowacki M, Nowak JK, Plattner H, Poulain J, Ruiz F, Serrano V, Zagulski M, Dessen P, Bétermier M, Weissenbach J, Scarpelli C, Schächter V, Sperling L, Meyer E, Cohen J, Wincker P.** 2006. Global trends of whole genome duplications revealed by the ciliate *Paramecium tetraurelia*. Nature **444:**171–178. http://dx.doi.org/10.1038/nature05230.
37. **Sperling L.** 2011. Remembrance of things past retrieved from the *Paramecium* genome. Res. Microbiol. **162:**587–597. http://dx.doi.org/10.1016/j.resmic.2011.02.012.
38. **Duret L, Cohen J, Jubin C, Dessen P, Goût JF, Mousset S, Aury JM, Jaillon O, Noël B, Arnaiz O, Bétermier M, Wincker P, Meyer E, Sperling L.** 2008. Analysis of sequence variability in the macronuclear DNA of *Paramecium tetraurelia*: a somatic view of the germline. Genome Res. **18:**585–596. http://dx.doi.org/10.1101/gr.074534.107.
39. **Forney JD, Blackburn EH.** 1988. Developmentally controlled telomere addition in wild-type and mutant paramecia. Mol. Cell. Biol. **8:**251–258.
40. **Phan HL, Forney J, Blackburn EH.** 1989. Analysis of *Paramecium* ma-

cronuclear DNA using pulsed field gel electrophoresis. J. Protozool. **36:** 402–408. http://dx.doi.org/10.1111/j.1550-7408.1989.tb05535.x.

41. **Simon MC, Marker S, Schmidt HJ.** 2006. Posttranscriptional control is a strong factor enabling exclusive expression of surface antigens in *Paramecium tetraurelia*. Gene Expr. **13:**167–178. http://dx.doi.org/10.3727/000000006783991809.
42. **Preer JR, Preer LB, Rudman BM.** 1981. mRNAs for the immobilization antigens of *Paramecium*. Proc. Natl. Acad. Sci. U. S. A. **78:**6776–6778. http://dx.doi.org/10.1073/pnas.78.11.6776.
43. **Jaillon O, Bouhouche K, Gout JF, Aury JM, Noel B, Saudemont B, Nowacki M, Serrano V, Porcel BM, Ségurens B, Le Mouël A, Lepère G, Schächter V, Bétermier M, Cohen J, Wincker P, Sperling L, Duret L, Meyer E.** 2008. Translational control of intron splicing in eukaryotes. Nature **451:**359–362. http://dx.doi.org/10.1038/nature06495.
44. **Duraisingh MT, Voss TS, Marty AJ, Duffy MF, Good RT, Thompson JK, Freitas-Junior LH, Scherf A, Crabb BS, Cowman AF.** 2005. Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. Cell **121:**13–24. http://dx.doi.org/10.1016/j.cell.2005.01.036.
45. **Curtenaz S, Beisson J.** 1996. In situ hybridisation as a method to study the regulation of gene expression in *Paramecium*. J. Eukaryot. Microbiol. **43:** 202–212. http://dx.doi.org/10.1111/j.1550-7408.1996.tb01392.x.
46. **Leeck CL, Forney JD.** 1996. The 5′ coding region of *Paramecium* surface antigen genes controls mutually exclusive transcription. Proc. Natl. Acad. Sci. U. S. A. **93:**2838–2843. http://dx.doi.org/10.1073/pnas.93.7.2838.
47. **Simmer F, Buscaino A, Kos-Braun IC, Kagansky A, Boukaba A, Urano T, Kerr AR, Allshire RC.** 2010. Hairpin RNA induces secondary small interfering RNA synthesis and silencing in *trans* in fission yeast. EMBO Rep. **11:**112–118. http://dx.doi.org/10.1038/embor.2009.273.
48. **Josse T, Teysset L, Todeschini AL, Sidor CM, Anxolabéhère D, Ronsseray S.** 2007. Telomeric *trans*-silencing: an epigenetic repression combining RNA silencing and heterochromatin formation. PLoS Genet. **3:**1633–1643. http://dx.doi.org/10.1071/journal.pgen.0030158.
49. **Todeschini AL, Teysset L, Delmarre V, Ronsseray S.** 2010. The epigenetic *trans*-silencing effect in *Drosophila* involves maternally transmitted small RNAs whose production depends on the piRNA pathway and HP1. PLoS One **5:**e11032. http://dx.doi.org/10.1371/journal.pone.0011032.
50. **Chen S, Krinsky BH, Long Y.** 2013. New genes as drivers of phenotypic evolution. Nat. Rev. Genet. **14:**645–660. http://dx.doi.org/10.1038/nri.2662.
51. **Capdeville Y.** 1971. Allelic modulation in *Paramecium aureliae* heterozygotes. Study of G serotypes in syngen 1. Mol. Gen. Genet. **112:**306–316. http://dx.doi.org/10.1007/BF00334432.
52. **Baudry C, Malinsky S, Restituito M, Kapusta A, Rosa S, Meyer E, Bétermier M.** 2009. PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. Genes Dev. **23:**2478–2483. http://dx.doi.org/10.1101/gad.547309.
53. **Arnaiz O, Mathy N, Baudry C, Malinsky S, Aury JM, Wilkes CD, Garnier O, Labadie K, Lauderdale BE, Le Mouël A, Marmignon A, Nowacki M, Poulain J, Prajer M, Wincker P, Meyer E, Duharcourt S, Duret L, Bétermier M, Sperling L.** 2012. The *Paramecium* germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. PLoS Genet. **8:**e1002984. http://dx.doi.org/10.1371/journal.pgen.1002984.
54. **Caron F.** 1992. A high degree of macronuclear chromosome polymorphism is generated by variable DNA rearrangements in *Paramecium primaurelia* during macronuclear differentiation. J. Mol. Biol. **225:**661–678. http://dx.doi.org/10.1016/0022-2836(92)90393-X.
55. **Le Mouël BA, Caron F, Meyer E.** 2003. Developmentally regulated chromosome fragmentation linked to imprecise elimination of repeated sequences in paramecia. Eukaryot. Cell **2:**1076–1090. http://dx.doi.org/10.1128/EC.2.5.1076-1090.2003.
56. **Garnier O, Serrano V, Duharcourt S, Meyer E.** 2004. RNA-mediated programming of developmental genome rearrangements in *Paramecium tetraurelia*. Mol. Cell. Biol. **24:**7370–7379. http://dx.doi.org/10.1128/MCB.24.17.7370-7379.2004.
57. **Pays E, Van Assel S, Laurent M, Darville M, Vervoort T, Van Meirvenne N, Steinert M.** 1983. Gene conversion as a mechanism for antigenic variation in trypanosomes. Cell **34:**371–381. http://dx.doi.org/10.1016/0092-8674(83)90371-9.
58. **Schildkraut E, Miller CA, Nickoloff JA.** 2005. Gene conversion and deletion frequencies during double-strand break repair in human cells are controlled by the distance between direct repeats. Nucleic Acids Res. **33:** 1574–1580. http://dx.doi.org/10.1093/nar/gki295.
59. **Galvani A, Sperling L.** 2002. RNA interference by feeding in *Paramecium*. Trends Genet. **18:**11–12. http://dx.doi.org/10.1016/S0168-9525(01)02548-3.
60. **Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD.** 2012. The Pfam protein families database. Nucleic Acids Res. **40:**D290–D301. http://dx.doi.org/10.1093/nar/gkr1065.
61. **Prat A, Katinka M, Caron F, Meyer E.** 1986. Nucleotide sequence of the *Paramecium primaurelia* G surface protein—a huge protein with a highly periodic structure. J. Mol. Biol. **189:**47–60. http://dx.doi.org/10.1016/0022-2836(86)90380-3.
62. **Benson G.** 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. **27:**573–580. http://dx.doi.org/10.1093/nar/27.2.573.
63. **Bendtsen JD, Nielsen H, von Heijne G, Brunak S.** 2004. Improved prediction of signal peptides: SignalP. J. Mol. Biol. **30:**340:783–795. http://dx.doi.org/10.1016/j.jmb.2004.05.028.
64. **Frankhauser N, Mäser P.** 1999. Identification of GPI anchor attachment signal by a Kohonen self-organization map. Bioinformatics **21:** 1846–1852.
65. **Krogh A, Larsson B, von Heijne G, Sonnhammer EL.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:**567–580. http://dx.doi.org/10.1006/jmbi.2000.4315.
66. **Nix DA, Eisen MB.** 2005. GATA: a graphic alignment tool for comparative sequence analysis. BMC Bioinformatics **6:**9. http://dx.doi.org/10.1186/1471-2105-6-S1-S9.
67. **Tamura K, Dudley J, Nei M, Kumar S.** 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24:**1596–1599. http://dx.doi.org/10.1093/molbev/msm092.
68. **Pfaffl MW.** 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. **29:**e45. http://dx.doi.org/10.1093/nar/29.9.e45.
69. **Arnaiz O, Goût J-F, Bétermier M, Bouhouche K, Cohen J, Duret L, Kapusta A, Meyer E, Sperling L.** 2010. Gene expression in a paleopolyploid: a transcriptome resource for the ciliate Paramecium tetraurelia. BMC Genomics **11:**547. http://dx.doi.org/10.1186/1471-2164-11-547/.
70. **Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP.** 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics **4:**249–264. http://dx.doi.org/10.1093/biostatistics/4.2.249.
71. **Edgar R, Domrachev M, Lash AE.** 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. **30:**207–210.
72. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods **9:**357–U354. http://dx.doi.org/10.1038/nmeth.1923.
73. **Sharp PM, Li WH.** 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15:**1281–1295. http://dx.doi.org/10.1093/nar/15.3.1281.
74. **Sehring IM, Mansfeld J, Reiner C, Wagner E, Plattner H, Kissmehl R.** 2007. The actin multigene family of *Paramecium tetraurelia*. BMC Genomics **8:**82. http://dx.doi.org/10.1186/1471-2164-8-82.